DOCUMENT RESUME

ED 122 326 CS 501 357

AUTHOR Addington, David W., Ed.; Kepke, Allen N., Ed.
TITLE Empirical Research in Theatre, Vol. 1.
INSTITUTION Bowling Green State Univ., Ohio. Center for Communications Research.
PUB DATE 71
NOTE 63p.
AVAILABLE FROM Center for Communications Research, Bowling Green State University, Bowling Green, Ohio 43403 ($2.00 prepaid)

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage
DESCRIPTORS *Communication (Thought Transfer); *Creativity Research; *Dramatics; Identification (Psychological); Listening Comprehension; Memorizing; Personality Studies; Readers Theater; *Speech Curriculum; *Theater Arts
IDENTIFIERS *Speech Communication

ABSTRACT

This journal provides a focal point for the collection and distribution of systematically processed information about theory and practice in theatre. Part of an irregularly published series, this issue contains investigations of the effect of quantity of stage lighting on audience comprehension, three methods of line memorization, reliability of judgments of directing techniques, actor-character personality identification in a theatre production, psychology and drama, and the effectiveness of reader's theatre and silent reading in communicating the meaning of unfamiliar dramatic literature. Research reports describe a correlation of personality types with level of approval of controversial material on the stage, and an analysis of the relationship between preference for specific set design styles and degree of introversion/extroversion. (KS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED122326

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EMPIRICAL RESEARCH IN THEATRE

Published by

The Center for Communications Research
Bowling Green State University

Volume I — Summer, 1971 — No. 1

The Effect of Quantity of Stage Lighting on Audience Comprehension 1

L. B. Rabby and L. S. Harms

When Actors Memorize Lines: An Examination of Three Methods 15

Richard Wuertz

Reliability of Judgments of Directing Techniques 22

Allen N. Kepke, John C. Stockwell, and Robert Zyromski

Actor-Character Personality Identification in a Theatre Production 29

R. Wayne Smith

Psychology and Drama 38

Brian Sutton-Smith and Gil Lazier

The Effectiveness of Reader's Theatre and Silent Reading in Communicating the Meaning of Unfamiliar Dramatic Literature 47

Howard D. Doll

RESEARCH REPORTS

Personality and Controversial Material on the Stage 56

Christopher M. Reynolds

Style and Temperament 58

Warren Pickett

Copyright 1971 by the Center for Communications Research

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Center for Communications Research

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

CS 501 357

EDITORIAL COMMENT: As many of you have probably noticed, the name of the monograph has been changed even before its first publication. The change was in the direction of parsimony, but the change was also appropriately symbolic. It represents the kinds of trauma which have accompanied the birth of this publication. We have indeed found experience to be a cruel teacher. We have made many senseless blunders and costly mistakes, and it appears from this vantage point, at least, that the job was far more difficult than we ever imagined possible. Often those difficulties resulted in reduced administrative efficiency, slow-downs, and mistakes. To those of you who have been tolerant of our errors and patient with our tardiness, especially our contributors, we would like to extend our deepest gratitude as we would, also, to those of you who have offered us many kind words of encouragement. We would also like to express our appreciation to our associate and assistant editors who have worked so generously to make this monograph possible.

Despite all the difficulties and effort, Empirical Research in Theatre is now more than simply an idea, it is manifest and in print, and we are candidly proud of it. At the same time, however, we do face some serious problems. This issue is really little more than a "trial balloon," or an experiment for which we must now await the "results." Bowling Green State University was most generous in supporting, entirely, this first issue. Nonetheless, no arrangements were made for supporting any subsequent issues. To put it bluntly, the experiment will be found successful and the journal will continue only if you demonstrate a need for this kind of publication by: 1) paying for copies, 2) asking for subsequent issues, 3) offering suggestions for improvement or support, or probably most important of all, 4) sending us articles. Please help us by responding to the last page of the journal. Your help and encouragement is both appreciated and necessary.

THE EFFECT OF QUANTITY OF STAGE LIGHTING ON AUDIENCE COMPREHENSION

L. B. Rabby and L. S. Harms

In an article in the Educational Theatre Journal, Jed Davis (1961) contends that theatre personnel "must recognize that their playwrights, directors, actors and designers have operated too long on cliches and speculation and intuition." Davis' specific reference was to children's theatre workers. However, his remarks are general enough to apply to the entire academic theatre program. Davis further states: "It is evident, however, that much more experimental study is needed on which to found the creative studies of the theatre."

As methodology for general experimental study becomes more refined, applications can be made to the problems unique in theatre production. These applications can afford us a firm basis on which we may attain evidence to support our intuitions and indicate to what degree our speculations are reliable and valid. This report is based on five related experimental research studies which bear on the problem of light levels and comprehension in theatre.

The amount of stage light in which an actor speaks affects audience comprehension of dialogue. Comprehension is also affected by the physical distance between an actor and audience members. The five studies reported here attempt to determine under experimental and performance conditions the direction and extent of these relationships.

The exact range of light level which lies between the extremes of darkness and glare has not been determined precisely for theatre production. McCandless (1958) suggests a range of 5 to 50 foot-candles; Selden and Sellman (1959) indicate 2 to 42 foot-candles as acceptable.

Authorities indicate that the lowest quantity of illumination used in stage lighting should not be so low as to cause the audience to strain in order to see comfortably, and the highest quantity should not be so high as to cause glare. It is suggested that the higher ranges of illumination have caused the audience to be more alert and have tended to make the production more enjoyable. Rowe (1960) points out, "compromises are compelled between lighting for mood and meaning and projection of effective expression of the actors to the reaches of the theatre auditorium."

It is likely that the effect of quantity of light on speech is relative to the contributions of the oral and visual channels of communication to comprehension. Evidence of the nature of the relationship between the two channels is provided by O'Neill (1954) in his demonstration that "vision contributed 44.5% to the understanding of vowels, 72% for consonants, 64.1% for words, and 25.9% for phrases."

Hollowell Davis (1960) summarizes the results of experimental research with this observation:

> In most conversation we employ both sight and hearing. Hearing is so predominant for most of us, however, that we seldom appreciate the importance of sight. The information that we get by watching not only the movement of the speaker's lips, but also his gesture and the expres-

sion of his face, is realized only when we hear with difficulty and are unable to see the speaker to our satisfaction.

On the relationship of distance to comprehension, Black (1957) has shown generally that intelligibility decreases as distance between speaker and listener increases. In the theatre setting, Harms and Whitaker (1963) have demonstrated that stage intelligibility decreases regularly as the distance between the actor and the audience member increases.

From an overview of the evidence stated above, the general hypotheses for this set of experimental studies were formed. These hypotheses were centered on the effect of quantity of stage lighting and distance of audience members on the comprehension of dialogue.

Two general hypotheses have been stated:

(1) Audience comprehension of the dialog of actors will vary proportionately with quantity of stage lighting used. In a low quantity of light, audience comprehension will be low, and as the light increases, comprehension will also increase.

(2) In any given quantity of light, comprehension will decrease as the distance between the actor and audience member increases, and increase as the distance decreases.

Cloze Procedure--"a psychological tool for gauging the degree of total correspondence between (1) the encoding habits of transmitters and (2) the decoding habits of receivers" (Taylor, 1954)--was employed in all experiments to measure audience comprehension.

Figure 1 contains a sample Cloze Procedure form. Note that every fifth

M: You (1)_____, too! You' lying and (2)_____ know it. You did, (3)_____! And there I was (4)_____ at that time lying (5)_____ bed not able to (6)_____ wondering how I was (7)_____ going to see you (8)_____ and crying my eyes (9)_____, while you--! I hate (10)_____! I wish you were (11)_____! I'm going home this (12)_____! I never want to (13)_____ eyes on you again! (14)_____ this time I mean (15)_____!
R: Muriel! Wait! Listen!
M: I (16)_____ want to listen! Let (17)_____ go! If you don't (18)_____ bite your hand!
R: I (19)_____ let you go! You've (20)_____ to let me explain! (21)_____ never--! Ouch! All right! (22)_____ if you want to--(23)_____ you haven't the decency (24)_____ let me explain! I (25)_____ you, too! I'll go and see Bell!

Figure 1. Sample Cloze Procedure form for Ah, Wilderness!

word is deleted and replaced by a ten-space line. The audience member first watched the scene under a particular lighting condition and was then given a form

and asked to reconstruct what the actor said by filling in the missing words. In scoring the forms, only the precise word from the play was counted as a correct response.

Experiment I

Procedure

The first of the five experiments was designed to establish the methodology to be used throughout the studies. This experiment involved three audiences, three scenes, and three levels of light. Each audience observed all three scenes, one in each of the different light levels.

A Greco-Latin rotation design, Figure 2, was employed to distribute

Sequence	Audience 1	Audience 2	Audience 3
1st	Scene 1 Light - High	Scene 2 Light - Low	Scene 3 Light - Mid
2nd	Scene 2 Light - Mid	Scene 3 Light - High	Scene 1 Light - Low
3rd	Scene 3 Light - Low	Scene 1 Light - Mid	Scene 2 Light - High

Sequence: The order of presentation to the three audiences.

Scenes:
1 - Tea and Sympathy scene.
2 - Ah, Wilderness! scene.
3 - The Diary of Anne Frank scene.

Figure 2. Greco-Latin rotation of scenes and light level.

equally the effects of audien e learning and fatigue, development of skill by the actors, and other extraneous variables which might have influenced the results of the experiment. This distribution permitted variables to act equally on each scene before each audience. Utilization of this rotation pattern allowed each scene to be presented to each audience in no more than one quantity of light and never in the same sequence.

The audiences were seated in four rows, equally spaced from the front

to the back of the auditorium. After each scene, a Cloze Procedure test was given to measure audience comprehension.

The University Theatre at Louisiana State University, a traditional proscenium arch theatre, built in 1931, was used for this experiment. Although the auditorium seats 574 persons, only selected seats on the main floor, containing 20 rows of 23 seats each, were actually used in the experiment. Specifically, the experimental audience was seated only in the center section in rows 2, 8, 13, and 19. The theatre was cooled to 72° and contained 40% relative humidity. During the experiment air conditioning fans were turned off (temperature was unchanged) and an attempt was made to reduce disturbing noises from outside and inside the theatre. No noise distractions were noticed during any of the experiments.

The three scenes selected for this experiment each had two characters, was suited to all three quantities of light, and could be performed in four to five minutes. The three scenes were: (1) Ah, Wilderness, Act 4, Scene 2, the scene between Muriel and Richard, (2) The Diary of Anne Frank, Act 2, Scene 1, the scene between Anne and Peter, and (3) Tea and Sympathy, the final scene of the play, between Laura and Tom.

The actor and actress selected for these three scenes were undergraduates, and were among the more experienced acting personnel available. Their articulation, enunciation, and projection were considered excellent.

The experimental design called for the selection of three quantities of stage lighting: approximately the highest level that would normally be used on the stage, the lowest level normally used, and the measured mid-point between the high and low levels. Based on recommendations of McCandless and Selden and Sellman, the following light levels were selected: (1) the low level, 2 foot-candles, (2) the mid level, 22 foot-candles, and, (3) the high level, 42 foot-candles. One light level was used in each scene with no variations in lighting during the scenes. No reflected light was measurable from the set or the background. The light levels were measured with a cosine-corrected foot-candle meter.

A light plot was designed to obtain an even dispersement of light over the acting area. This one light plot was used for all three scenes. Each subdivided area of the stage was cross-lighted from the bridge, beam, and tormentor positions. The same colors of gelatins were used for each scene: Brigham, number 17, Light Blue Lavender, from stage right, and Brigham, number 62, Salmon, from stage left. As the actors moved on stage, there were no noticeable "dead spots" or "hot spots."

The consistency of the actor's sound level was checked by use of a B&K instruments sound level meter, and the consistency of their loudness was evaluated subjectively by an experienced theatre director seated in the last row of the orchestra. When the experiments were completed the experimenter was assured that the levels had, indeed, remained relatively consistent throughout all of the experimental conditions.

A minimal amount of scenery was used for the scenes. All three were played against a set of black velour curtains. The furnishings were arranged to keep the movement of the actors in the same stage areas and levels for each scene,

and blocking was designed to keep the actors in shared positions and on the same stage levels. These positions were not considered critical, however, since the blocking for all three scenes before all three audiences remained the same. While blocking and rehearsing these scenes, an attempt was made to focus attention on the speaking actor in order to attain maximum comprehension. During the experiment, a careful check was made to see that the blocking remained the same.

With the exception of powder, no makeup was used for this experiment. The costumes consisted of street clothing which was kept in the range of dark values of colors. The costumes were changed for each of the three scenes but were the same for each audience for any given scene.

The audiences for this experiment were composed of Speech Fundamentals and beginning Public Address students. A minimum of 40 students were in each audience, which allowed more than ten students to sit in each of the four rows used. On entering the auditorium, the participants were given a seating placement card. Brief standardized instructions were read to each audience by the experimenter. The audiences saw each scene, which lasted approximately five minutes, then took the Cloze Procedure test which lasted approximately ten minutes. Each audience member was required to make 25 responses to each scene.

In order to equalize the cell n's, ten Cloze Procedure tests were selected randomly from among those obtained from each row for statistical analysis.

Two-way analysis of variance with a 3 x 4 split-plot design was used to test the data. The two main dimensions of the design were: (1) light--three variations, and (2) distance--four levels according to the row in which the audience member was seated. The split-plot design allowed for tests of linearity of the two main effects by testing the cubic and quadratic effects. The .05 level of significance was considered critical in all statistical tests.

Results

The results of that analysis are displayed in Table 1. It is apparent from Table 1 that the interaction of the two main variables (light and distance) was not significant. Evidently comprehension is not differentially effected by the interaction of light level with distance from the stage. Or more simply, the effects of level and distance on comprehension are relatively independent.

With the interaction established as non-significant, it was then possible to consider the significance of the main effects. It is also apparent from Table 1 that both of the main effects were significant, suggesting that, indeed, comprehension was affected by variations in light level, and that similarly those scores were affected by differences in the distance between the audience member and the stage. The analysis indicated further (see Table 1) that deviations from linearity from either light level or distance from the stage as measured by quadratic or cubic effect were not statistically significant forcing us to reject the hypothesis that the effects of either light level or distance on comprehension were other than linear.

To further illustrate the linearity of the effects, those effects were graphed in Figures 3 and 4. Figure 3 shows the effects of variations in light level on the comprehension scores, while Figure 4 shows the effects of distance.

TABLE I

ANALYSIS OF VARIANCE: EXPERIMENT I

Source of Variance	Sums of Squares	d.f.	Mean Squares	F
Light Level	76.00	2	38.00	4.897*
Linear	73.70	1	73.70	9.500
Quadratic	3.30	1	3.30	1.000
Distance	213.00	3	71.00	9.149*
Linear	209.44	1	209.44	26.970
Quadratic	0	1	0	0
Cubic	5.33	1	5.33	1.000
Level x Distance	11.00	6	1.83	.234
Error	2701.00	348	7.76	
Total	3001.00	359		

*Significant at .05 level of confidence.

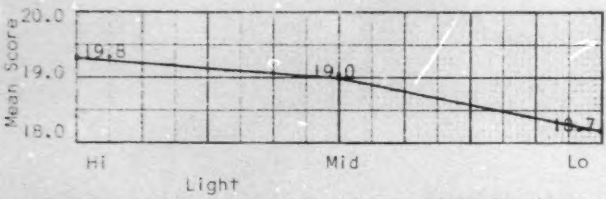

Figure 3. Experiment I--Comprehension by Light level.

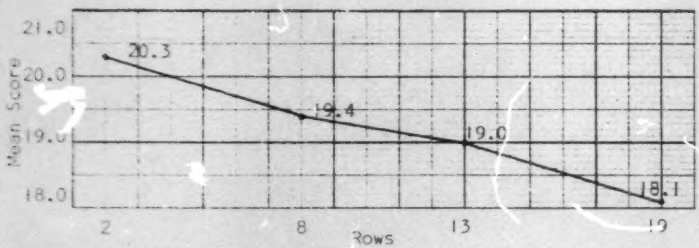

Figure 4. Experiment I--Comprehension by distance.

Summary

The results of this experiment indicated comprehension varied proportionately with the quantity of stage lighting and that the progression of effect was linear. The smaller the quantity of light the lower the comprehension score, and the greater the quantity, the higher the score.

Likewise it was apparent that comprehension varied inversely with the distance from the audience to the stage and that that variation was also linear. The greater the distance between the audience and the stage the lower the comprehension score, and the less the distance the higher the score.

It was revealed, furthermore, that the effects on comprehension due to lighting level did not interact with the effects due to distance from the stage. Apparently the effects of one are relatively independent of the effects of the other.

Experiment II

Procedure

The second study involved further testing the results of the first experiment. If comprehension acted directly as a function of one unchanged level of light used throughout a scene, then the level of lighting could be changed during the scene and an appropriate change of comprehension should occur.

The hypothesis for this second experiment was:

> Audience comprehension of dialogue of actors in changing quantities of stage lighting will vary in direct relationship to the direction and amount of the change. If stage lighting moves from a low to a high level during a scene, comprehension will increase. Conversely, if stage lighting is changed from a high to a low level during a scene, comprehension will decrease.

The design and procedures of this second experiment were similar to the first. Two scenes from Cyrano de Bergerac were used: the meeting of Cyrano and Roxanne in Ragueneau's pastry shop in Act II, and part of the balcony scene between Cyrano and Roxanne in Act III. The two scenes were performed in an experimental environment before one audience in chronological order.

A pattern of light level changes was designed to occur during each scene. Using the light levels from Experiment I, the lighting changed by steps from low to mid to high in the first scene. The sequence was reversed for the second scene. These three segments became the test segments.

The Cloze Procedure tests were arranged to permit sampling only during three test sections and to request 20 responses for each level of light in each scene.

The second experiment was performed in the University Theatre at the University of Kansas; a proscenium arch theatre, completed in 1957, with a seating

capacity of 1,188 persons. The replication of the experiments in different theatres was an intentional attempt to increase our ability to generalize from the results of the studies. For this experiment distance was not an independent variable, instead it was controlled simply by having the entire experimental audience sit on one row in the center of the auditorium.

The actors used in this study were those playing lead roles in Cyrano which was in public performance during the time of the experiment. The scenery, blocking, and light plot were those used in the public performance.

The experimental audience was composed of Speech Fundamentals students, and the procedure was similar to that of Experiment I, with the exception that only two scenes and one audience were used. V/U Meter readings of the actor's sound pressure levels suggested that no important changes took place during the experiment.

Twelve participants in this experiment made 20 responses for each level of light in each scene. The total number of responses per level of light was 480 with 20 as the highest score possible.

Two-way analysis of variance was once again used to test the data. This time the major dimensions were: (1) the scenes (two variations), and (2) lighting levels (three variations). Correlation coefficients were also computed to test the degree to which the comprehension scores were affected by variations in lighting level.

Results

The results of the analysis of variance, displayed in Table 2, reveal

TABLE 2

ANALYSIS OF VARIANCE: EXPERIMENT II
CYRANO DE BERGERAC

Source	Sums of Squares	d.f.	Mean Squares	F-Ratio
Scenes	.22	1	.22	.021
Light	236.11	2	131.56	13.03*
S x L	2.12	2	1.06	.105
Error	60.38	6	10.10	
Total	298.83	11		

*Significant at the .05 level of confidence.

no statistically significant F-ratios. Apparently comprehension was affected by neither the interaction of the lighting levels with the two different scenes, the differences in the two scenes, nor by the differences in the lighting levels, though the latter did approach significance. Nonetheless the correlation coef-

ficient of the relationship of the lighting variations with changes in comprehension was statistically significant. Those effects are represented graphically in Figure 5.

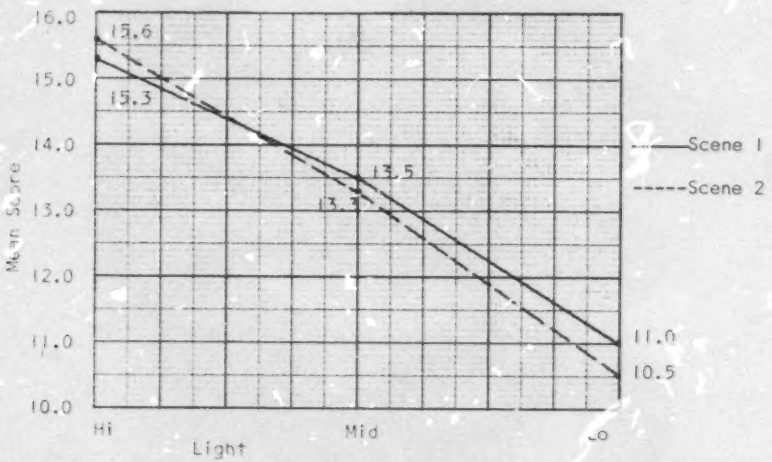

Figure 5. Experiment II--Comprehension by light level in Cyrano de Bergerac.

Summary

Although the findings from this experiment were inconsistent and therefore inconclusive, it is likely that variations in the quantity of lighting within a scene may be accompanied by significant shifts in comprehension. That is to say that, as the lighting is increased within a scene, it is likely that that part of the scene will be met with greater comprehension than will another part of the scene under less intense illumination.

Experiment III

Procedure

The third experiment in this series was a two-part replication designed to determine whether the results of the previous laboratory experiments would be validated in actual performance environments. Two scenes selected were from Mary Stuart which was in production at the time. In these two scenes, one from each act, the Lord of Leicester had long uninterrupted speeches. The light levels to those two scenes during performance were: Scene one in Elizabeth's court, 24 foot-candles, and scene two in the prison, 2 foot-candles. No lighting changes occurred during the scenes.

In order to test the audience comprehension of these two scenes under performance conditions, Cloze Procedure tests were given to a group of audience members who represented a randomly selected cross section of the normal theatre audience. At the end of Act I, this group was instructed to fill in Test 1, and at the end of the last act, they were asked to fill in Test 2.

In the second half of this experiment the same two scenes were presented to an experimental group. A system of rotation was employed; the scene order was reversed and the light levels for the scenes were reversed. No lighting change occurred during the experimental performance and all other conditions were the same as those in the actual public performance: same actors, blocking, light plot, light color, scenery, etc. A Cloze Procedure test was given to the experimental audience after each scene.

The theatre in which this study was conducted was the Experimental Theatre at the University of Kansas; a small, intimate theatre with a maximum distance of 45 feet between the audience and stage. The seating arrangement for both experimental and performance groups was randomly assigned, and no attempt was made either to control or to test the effects of distance. The experimental audience was composed of Speech Fundamentals and Theatre History students.

Identical tests were applied to both sets of data--experimental and performance. The relationship between the light levels and the comprehension scores was determined by computing bi-serial coefficients of correlation, and the statistical significance of the difference between the scenes was tested using the t statistic.

Results

The results of all of the statistical tests may be found displayed in Table 3. It is apparent that both correlations are relatively high and statisti-

TABLE 3

TESTS OF EXPERIMENTAL AND PERFORMANCE MEASUREMENTS

Performance Conditions		Experimental Conditions	
$\bar{X}$, Scene 1 @ 24 f.c. = 15.22		$\bar{X}$, Scene 2 @ 24 f.c. = 14.93	
$\bar{X}$, Scene 2 @ .2 f.c. = 14.36		$\bar{X}$, Scene 1 @ 24 f.c. = 12.29	
r = .55*	t = 3.28	r = .54*	t = 1.56*

*Significant at the .05 level of confidence.
Foot-candles are abbreviated f.c.

cally significant indicating that as the quantity of stage lighting increases there is a subsequent increase in comprehension regardless of whether the scene was viewed in an experimental or normal performance setting. As a rule of thumb one may square correlation coefficients to obtain the amount of variance in one

variable attributable to variations in another variable, and using that rule of thumb, we may tentatively conclude that 25% of the variance in comprehension scores may be attributable to changes in quantity of light on the stage. Nonetheless, some degree of caution should be exercised when we generalize from the results of this phase of the study since the t-test of the difference between the comprehension scores for the two lighting conditions for the performance audience was not found to be statistically significant.

Summary

This third study investigated the performance of two scenes under experimental and performance conditions. This investigation revealed findings similar to studies one and two and provided additional support to the study's major hypothesis; that comprehension scores vary proportionately with the amount of light on the stage. The findings of this phase extend the generalization to include more intimate theatres and audiences in a more natural setting.

Experiment IV

Procedure

The fourth experiment, based on information obtained in the performance part of Experiment III, was designed to measure the comprehension of dialogue from two scenes observed under performance conditions in two different quantities of light at four distances from the stage. The stimulus for this study was from a musical production, Paint Your Wagon, in public performance at the time of the experiment. The experimental scene was selected from each act in which the same actress had dialogue appropriate for testing. The two scenes selected were played in two extreme levels of lighting: 46 foot-candles for the high level, and 4 foot-candles for the low level.

Performances of the musical were attended by Speech Fundamentals students who were equally distanced from front to back on rows 1, 7, 14, and 21 in The University Theatre at the University of Kansas. Cloze Procedure tests were administered to the subjects at the end of each act.

Results

Statistical tests of the results of this part of the experiment were not found to be statistically significant even though a trend in the direction of the results may seem apparent in Figure 6. The n for this part of the test was relatively small and may have partially accounted for the failure to attain statistical significance.

Experiment V

Procedure

Experiment V was the last in the series, and it checked comprehension once again under performance conditions. This time the stimulus was The Cherry Orchard performed in The University Theatre at the University of Kansas. The 13 subjects used in this phase of the study were seated randomly on the lower

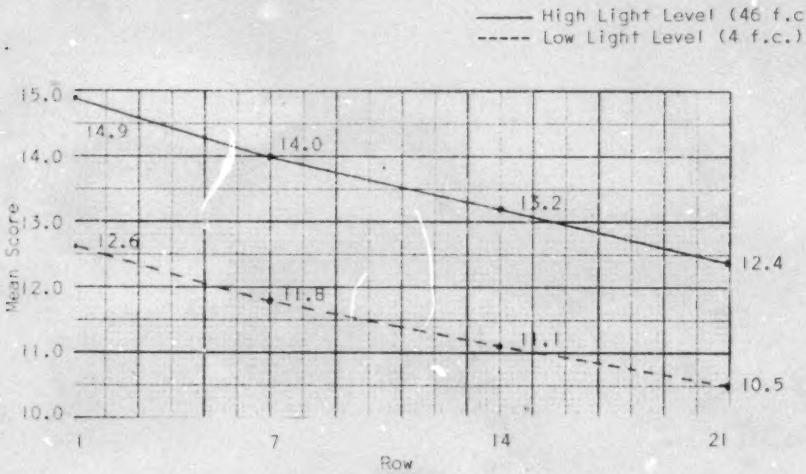

Figure 6. Experiment IV--Comprehension by light level and distance in Paint Your Wagon.

floor of the auditorium. Each subject was requested to make 30 Cloze responses to each section of the play. The two test sections were taken from different acts and both were taken from long speeches delivered by Yermolay. All of the blocking, lighting, costumes, and make up were those used in the public performances of the play. In all other respects the methodology was similar to that of Experiment IV, except that the present study included a far narrower range of lighting variation; low level lighting was 14 foot-candles, and high level lighting was 22 foot-candles.

Again the data were tested using bi-serial correlation and t-tests.

Results

The results of Experiment V are displayed in Table 4.

TABLE 4

TESTS OF PERFORMANCE MEASURES: THE CHERRY ORCHARD

Scene 1 @ 22 f.c.	Scene 2 @ 14 f.c.	r	t
n = 13, $\overline{X}$ = 20.69	n = 13, $\overline{X}$ = 17.15	.65*	1.70

*Significant at .05 level of confidence.

Summary

Once again the results are somewhat inconclusive. The correlation coefficient (Table 4) suggests a strong and significant relationship between the quantity of light and comprehension, but that relationship is not supported by a statistically significant difference between the mean scores for the two scenes.

It is important to point out here that only 8 foot-candles separated the two scenes, and that lack of power may well account for our failure to detect statistical significance when the difference between the two scenes was tested.

General Conclusions

At this stage in the research on the effect of quantity of stage lighting on audience comprehension, certain general conclusions may be outlined:

1. The amount of stage light affects audiences' comprehension of actor's dialogue.

2. As expected, comprehension decreases as the distance between the actor and the stage increases.

3. Comprehension scores do not seem to be affected by the interaction of distance from the stage and level of intensity of stage lighting.

4. The results seem to become less conclusive when tests are shifted from a more controlled laboratory environment to a more natural performance setting.

The results of the studies suggest that correlation statistics may be the more appropriate statistic for analyzing data of this sort, since we are more interested in common variances than we are in differences between means. That any of the means were significantly different is perhaps more surprising than the fact that all of the correlation coefficients were statistically significant.

Future research might be conducted with only correlations in mind and thus they might explore a greater number of lighting variations with greater precision. Whatever the statistic, research is needed to map the perameters of the relationship of this psychophysical phenomena; where do we reach the points of diminishing returns, and at what point in lighting intensity do we generally attain the highest comprehension scores.

The experiments reported here have dealt primarily with quantity of light and secondarily with distance. The effects of other variables such as stage position, stage levels, stage planes, costumes, sets, and the color of the lighting on comprehension might also bear fruitful investigation.

REFERENCES

Dr. Rabby is Director of Theatre at Ohio Wesleyan University, and Dr. Harms is a Professor of Speech Communication at the University of Hawaii.

Black, J. W. Multiple-Choice Intelligibility Tests. Journal of Speech and Hearing Disorders, 1957, 22, 213-235.

Davis, H. (Ed.) Hearing and Deafness. New York, Holt, Rinehart & Winston, 1960, p. 354.

Davis, J. Prospectus for Research in Children's Theatre. Educational Theatre Journal, 1961, 13, 274-275.

Harms, L. S. & Whitaker, B. The Stage Intelligibility of Experienced and Inexperienced Student Actors and Actresses. Western Speech, 1963, 27, 25-29.

McCandless, S. A Method of Lighting the Stage. New York, Theatre Arts Books, 1958, p. 34.

O'Neill, J. J. Contributions of the Visual Components of Oral Symbols to Speech Comprehension. Journal of Speech and Hearing Disorders, 1954, 19, 439.

Rowe, K. T. A Theatre in Your Head. New York, Funk & Wagnalls, 1960, p. 76.

Selden, S., & Sellman, H. D. Stage Scenery and Lighting. New York, Appleton-Century-Crofts, 1959, p. 217.

Taylor, W. L. Application of "Cloze" and Entropy Measures to the Study of Contextral Constraint in Samples of Continuous Prose. Unpublished Doctoral Dissertation, University of Illinois, 1954, p. 3.

WHEN ACTORS MEMORIZE LINES: AN EXAMINATION OF THREE METHODS

Richard Wuertz

Most actors must memorize lines. Usually, this means an actor is expected to recall not only the ideas but the exact words of his character as shown in the script. For some actors, memorizing lines is difficult, and the actor who has trouble with it is a frustration to the other actors, to the director, and sometimes to the audience. The present inquiry was directed to the question: is there a method of memorizing lines that would be most helpful to most actors?

The traditional literature in acting and directing reveals a variety of memorization approaches and a variety of opinions regarding their relative values (Albright, 1959, p. 128; Blunt, 1966, p. 128; Covan, 1940, p. 246; Dolman, 1949, p. 78; Eustis, 1937, pp. 30-97; Goodman, 1956, pp. 190-191; Kahan, 1962, p. 128; McGaw, 1955, p. 135; Muni, 1949, p. 528). From these writings, three methods were discerned: 1) using either whole cues or "sides," an actor concentrates on repeating the words as "sounds" rather than as meaningful utterances--what might be called "Rote"; 2) using whole cues and understanding the underlying ideas of the word groups, an actor repeats the words as meaningful utterances--what might be called "Meaningful"; and 3) using whole cues, understanding the ideas, and possessing a concept of his character, an actor repeats the words as meaningful utterances and uses characteristic movement--what might be called "Characterization."

Although most relevant research in psychology has focused upon disconnected lexical items--sense and nonsense words and syllables--some conclusions regarding memorization can be drawn: 1) it is improbable that meaning can be avoided even when a "Rote" method is attempted; 2) different methods of memorizing tend to make material more or less meaningful for subjects; and 3) different methods of memorizing tend to result in different amounts of recall (Asher, 1966, p. 79; Ausubel, 1963, p. 22; Deese, 1961, p. 11; Deese & Hulse, 1967, pp. 254-256; Dowling & Braun, 1957, p. 216; King & Russell, 1966, p. 481; Marks & Jacks, 1952, Poulton & Brown, 1967; Woodhead, 1966).

The study described below was planned with an awareness of these conclusions. An attempt was made to consider different methods of memorization arising out of theatre practice and speculation. Further, an attempt was made to model the study after a practical situation in theatre. During the earlier stages of rehearsal, with the help of someone to "cue" them, actors may attempt the memorization of short scenes during short working periods in preparation for rehearsal. The study was designed as an attempt to answer the following question: For small amounts of material and short memorization periods, are there any significant differences in immediate and short-term recall among three methods of memorizing dramatic dialogue?

Subjects

Thirty secondary school student volunteers participated in the experiment. They were especially appropriate for the study because they are the novices to the theatre, with little or no experience in memorizing lines.

Thirty volunteers were gathered into a total group equal in number of males and females, and equal in number of students from each of the sophomore, junior, and senior classes. This total group was divided into three subgroups of ten each, maintaining approximately the same balance. Random assignment to the subgroups within sex and year-in-school was made.

Materials

Portions of three plays (Incident at Vichy, Miller, 1964, p. 7; Dumb Waiter, Pinter, 1960, pp. 86-88; When We Dead Awaken, Poulson, 1962, pp. 337-338) were chosen with the recognition that, since little is known about variables of dramatic dialogue, it was impossible to select three scripts which could be considered comparable. However, the design was expected to reveal, as statistical interaction, the tendency of a script to facilitate more efficient line memorization than other scripts. Therefore, the only requirements for selection were that the three excerpts use modern English, that the number of cue words and words to be memorized be held constant from script to script, and that the name of the character whose lines were to be learned be alterable to one that corresponds to the sex of the subject without violating the plausibility of the scene.

Procedure

The procedure consisted of training a subject to use a method, have him use it, test his recall immediately, and test again four hours later. For all methods some similarities existed. The actual time with the script trials was the same, 15 minutes. Before trials each subject was told that he could rest if he desired, that he could violate the order of the script and run through portions of it which were vague, and that he should try to break his dependency on the script when he began to feel secure about any line or lines. For all methods, a test immediately followed each trial under trial conditions except that a tape recorder was activated and the subject was not allowed the use of the script. For all methods, a retest followed the test four hours later. No scripts were made available to subjects between the test and retest periods.

Method: Rote (R)

The label for this condition was chosen because of its procedural similarity to conditions described in theatre literature where the tendency is to focus the subject's attention on the words, without supplying him with or encouraging him to adopt a meaningful context.

The subject was given a script that contained his lines preceded by the last word of each cue speech. During most of the trial period, the experimenter read aloud only the last word of each cue speech, and the subject attempted to repeat the appropriate response line. For the last run-through, the experimenter read the whole cue speech in order to better prepare the subject for the testing situation.

Method: Interpretation (I)

This method was designed to incorporate the suggestions by authors for memorizing in a meaningful way. An attempt was made to focus the subject's

attention first on the idea underlying the words and only later upon the exact words.

Unlike the Rote condition, this method required a preparation period before the subject actually worked with the script. During this period the subject was supplied with background information about the character and the situation considered relevant to the scene, but the actual script was withheld.

During the trial period, the time spent with the script was similar to the Rote method in that it allowed as many run-throughs of the script as possible during the period. However, it differed in some important ways. First, the subject was supplied with a script that had character names introducing each line, and entire lines were given--in effect, the script had the look of a normal excerpt from a play. In addition, after the subject read each line, he was asked to paraphrase it, and to continue using a paraphrase until he became familiar enough with the "idea" to recall it with no more than a quick glance at the script. After becoming familiar he was then to turn his attention to the exact words in the script and attempt to master them on subsequent run-throughs.

The essence of the "Interpretation" method, then, was for the subject to concentrate on ideas, paraphrase, and only gradually turn to the exact words.

Method: Characterization (C)

This method was modeled after the third method distinguished in acting texts. Characterization was similar to interpretation, but differed in that the subject was encouraged to attempt character creation and to move "in character" while memorizing lines.

Preceding the trial period was a preparation period that included supplying the subject with the kind of information given during the interpretation preparation period. In addition, the subject was given the character's primary or scene objective, and his specific objectives. The primary objective was a statement describing what the character was trying to accomplish in the scene. The specific objectives were statements describing minor goals or purposes that, if accomplished, would help the character to fulfill his larger purpose. The subject was then placed on a rehearsal set and given some place-to-place movements for his character. He then repeated the character's objectives and moved accordingly on the set. He had been trained to understand that words and movements are manifest attempts to accomplish the character's objectives. Consequently he was to try to use words and movements to accomplish the given objectives. When the subject approached the trial, he had the objectives and movements in mind; all that was missing were the words.

The subject was handed a script similar to that used for the interpretation method, but, in addition, it contained his character's movements. In order to encourage him to understand the relation between the lines and the objectives, he first wrote the objectives in the margin beside the lines to which they applied. Then, he went onto the rehearsal set, followed the character's movements, and used the interpretation paraphrase technique.

Dependent Measures

The raw data of the study were tape recordings of subjects' attempts to recall the lines of three experimental scripts, both immediately after the trial period (Test) and four hours later (Retest). "How well a line has been learned" seemed measurable three different ways. One way involved recapturing the general sense of the line ("idea estimation"). Here, three judges rated how accurately each subject came to the underlying "idea" of each group of words set off by a semicolon or period.

The second measure consisted of comparing recall responses to the wording of the script. This yielded a number corresponding to the number of correct words recalled from which was subtracted one-fifth the number of incorrect words recalled. A word was considered correct if it occurred in the same sequence as indicated in the script. It was believed that a word uttered in the recall session that was either out of sequence or in addition to the words in the script should count against the subject. However, simple subtraction of each incorrect word seemed too harsh a penalty, and the percentage of such words to be subtracted from the total correct--one-fifth--was chosen as an arbitrary estimate of an appropriate weighting of penalty.

The final measure was a stop watch count of hesitation time or "latency." Hesitations vocalized or silent, immediately after a cue, and during line recall were timed. A period of two seconds of silence was allowed immediately after a cue in order to avoid penalizing a subject for using silence as a communication device. This measure yielded a total latency time for each subject on each method. All three measures were employed in the analysis of data.

The two testing periods and the three measures yielded six matrices of data. The data were submitted to a Lindquist Type II analysis of variance design (Lindquist, 1956, pp. 273-281). The .05 level of confidence was chosen as the level at which to reject the hypothesis of no difference. Five of the six F-ratios were statistically significant and were also found to have statistical interaction. The presence of interaction between mode of memorization and experimental script on the other five tests necessitated reporting any significant differences between one mode of memorization and another seperately for each of the scripts, since, no single mode of memorization was significantly different from another for all scripts. T-tests were run to reveal significant differences between memorization modes for each script within each of the five significant F-ratios. Again, the level of confidence was set at .05.

Results

The findings of the study support the conclusion that for small amounts of dramatic dialogue and short memorization periods, there are significant differences in immediate and short-term recall among three methods of memorization. However, there are numerous qualifications.

The script by mode interaction clearly reveals that no single mode of memorization is superior to another for all three scripts used in the study. While there are indications that the Rote method may be least effective for the

TABLE

SIGNIFICANT DIFFERENCES AMONG MODES
OF MEMORIZATION FOR ALL SCRIPTS AND ALL MEASURES*

	Dumb Waiter	Incident at Vichy	When We Dead Awaken
Idea Accuracy:			
Test			C>R; I>R
Retest			
Word Count:			
Test		R>I; R>C	
Retest	I>C	R>C	C>R
Latency:			
Test	I>R; C>R		
Retest	I>R; C>R		

*R = Rote, I = Interpretation, and C = Characterization. Only differences significant beyond the .05 level of confidence are recorded in the table.

excerpts from When We Dead Awaken and Dumb Waiter, there are similar indications that it is most effective for excerpts from Incident at Vichy (see Table). This suggests that the way dialogue is constructed in plays may have considerable influence upon the ease or difficulty of memorization and, depending upon characteristics as yet unknown, may be more efficient or appropriate with one memorization approach than another. The dialogue of the three experimental scripts was examined and a number of differences were noted. Only one seemed related to the findings: the length of the speeches to be memorized. The excerpt from Incident at Vichy contained speeches which were considerably longer than similar response-lines in the other two scripts. The findings for the Incident at Vichy script were also the only ones where Rote proved superior to the other modes of memorization. Obviously, since the absolute number of words to be memorized was held constant across scripts, the longer speeches of the Incident at Vichy script were confounded with other variables. Where words are held constant, longer speeches mean fewer response-lines to be learned and fewer cue-interruptions. However, the notion that a rote method might work particularly well in the memorization of extended, connected utterances forms a hypothesis with some face validity. Future studies in line memorization will undoubtedly give greater consideration to script variables; among which length of speeches may be promising.

The choice of three different criteria for measuring how well a line has been memorized provided information which would not otherwise have been available. Many of the significant differences between modes of memorization would not have emerged if only one criterion had been used, since the findings through any one measure are not repeated in any of the other measures. Further

work to establish the independence of these measures seems indicated. It appears that inquiry into the effectiveness of a memorization method may have to take into account how one chooses to define memorization, and that the choice may depend upon how significant such practical cues as amount of hesitation time, exact wording, and grasp of the underlying idea are to a director or researcher. In other words, one approach to memorizing lines may prove to be more effective than another when one kind of memorized result is desired but not when another kind is sought.

It may be observed that only in two instances were significant mode differences found on the immediate test repeated on the retest four hours later (see Table). Mode differences for the Dumb Waiter script, measured through latency, were found on the test and re-emerged on the retest. Similarly, for the Incident at Vichy script, one of the significant findings on the test, as measured by word count, was repeated on the retest. No other significant differences found on the test reappeared on the retest, and several significant differences emerged on the retest for the first time. There is a suggestion here that effect of a mode of memorization may not be stable over time, and further investigation seems warrented.

Finally, it may be well to emphasize the possible importance of subject differences. While this was not the focus of the present study, it is likely that an approach to memorizing lines may be better for some persons, while another approach may be better for others. Further studies along this line may be of considerable value.

In summary, then, the study reveals no "best method" of memorizing lines in plays. But it does suggest that one method may be better than another (1) for a particular kind of script, (2) for a particular kind of "memorized" result, and (3) when immediate or short-term retention is desired.

Reference

Mr. Richard Wuertz is an instructor of acting and speech at Marshalltown Community College, Marshalltown, Iowa. The research described in this article was the basis for a masters thesis conducted under the advice of Dr. George Gunkle at the University of Iowa, 1969.

Albright, H. D. Working Up A Part. Boston: Houghton Mifflin, 1959.

Asher, J. J. The Learning Strategy of the Total Physical Response: A Review. Modern Language Journal, 1966, 50, 79.

Ausubel, D. P. The Psychology of Meaningful Verbal Learning. New York: Grune and Stratton, 1963.

Blunt, J. The Composite Art of Acting. New York: The Macmillian Co., 1966.

Colvan, E. B. Face the Footlights. New York: McGraw-Hill, 1940.

Deese, J. From the Isolated Verbal Unit to Connected Discourse. Verbal Learning and Verbal Behavior. New York: McGraw-Hill, 1961.

Deese, J., Hulse, S. H. The Psychology of Learning, 3rd. ed. New York: McGraw-Hill, 1967.

Dolman, J. The Art of Acting. New York: Harper and Brothers, 1949.

Dowling, R. M., Braun, H. W. Retention and Meaningfulness of Material. Journal of Experimental Psychology, 1957, 54, 216.

Eustis, M. Players at Work. New York: Theatre Arts, 1937.

Goodman, E. Make Believe--The Art of Acting. New York: Charles Scribner's Sons, 1956.

Kahan, S. Introduction to Acting. New York: Harcourt, Brace and World, 1962.

King, D. J., Russell, G. W. A Comparison of Rote and Meaningful Learning Material. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 481.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin, 1956.

Marks, M. R., Jacks, O. "Verbal Context and Memory Span for Meaningful Material". The American Journal of Psychology, (1952), 65, 299-300.

McGaw, C. Acting is Believing. New York: Holt, Rinehart, and Winston, 1955.

Miller, A. Incident at Vichy. New York: Dramatists Play Service, 1964.

Muni, P. The Mechanics of Movie Acting. Actors on Acting. ed. Cole, T., Chinoy, H. K., New York: Crown Publishers, 1949.

Pinter, H. The Caretaker and the Dumb Waiter. New York: Grove Press, 1960.

Poulson, A., trans. Last Plays of Henrik Ibsen. New York: Bantam Books, 1962.

Poulton, E. C., Brown, C. H. Memory After Reading Aloud and Reading Silently. British Journal of Psychology, 1967, 58, 221.

Woodhead, M. M. Simple Rehearsal Strategies for Short-Term Recall. Psychonomic Science, 1966, 6, 385-386.

RELIABILITY OF JUDGMENTS OF DIRECTING TECHNIQUES

Allen N. Kepke, John C. Stockwell, Robert Zyromski

In educational theatre, one of the most common activities of teachers and directors is making evaluations of performances. Such judging activities assume that "expert faculty judges"--are able to make value judgments with some degree of reliability. According to a growing list of studies in both speech and drama, however, agreement among the "experts" is not as strong as we would like it to be. These studies have consistently revealed a need for improved evaluative techniques.

The major purpose of this study was to determine the reliability of ratings of performances directed by persons of varying ability as judged by "competent or expert" judges. A secondary purpose was to determine the reliability of several judging procedures.

The present research has precedent in research which studied the reliability of judgments of acting performances in the theatre (Clevenger & Tolch, 1962; Richardson & Waal, 1966). Although the phenomenon to be judged here is directing, the problem of consistency of evaluation is equally important.

Richardson and Waal, building on the efforts of Tolch and Clevenger concluded (1) that judges of acting performances could assign a rating to specific criteria--projection, timing, etc.--more reliably than they could assign a rating to a total performance; (2) that judges could rank performances more reliably than they could rate either specific criteria or total performance; and (3) that night-to-night shifts of intuitive notions of "good," "bad," and "average" could be reduced by specifying criteria. Though other conclusions emerged, those served as the bases on which the hypotheses of this study were developed.

H_1: Judges rate "specific criteria" of directing excellence more reliably than they rate "overall performance."

H_2: Judges rank directing performances more reliably than they rate either "specific criteria" or an "overall performance."

H_3: Night-to-night shifts of intuitive standards of judgment will be affected by specifying criteria.

Procedure

Eight graduate students and two faculty members from the Department of Speech at Bowling Green State University were assembled on two different evenings approximately two weeks apart to evaluate one-act plays directed by undergraduate students. Three different plays were presented each evening. The actors were all undergraduates with varying amounts of theatrical experience.

The critic-judges were two faculty members, seven doctoral fellows, and one masters student. All had received extensive training in theatre production and many of the doctoral students had taught for several years before returning for further graduate study.

Three different methods were used to evaluate the plays. (Copies of the questionnaire may be obtained from Dr. Allen N. Kepke, Speech Department, Bowling Green State University, Bowling Green, Ohio, 43403). First, the judges were asked to offer an "overall evaluation" of the total directing performance on a five-point equal interval rating scale ranging from "worst" to "best" possible. This response was elicited first to avoid "contamination" by the evaluations to follow. Secondly, the judges were given 5 questions, each of which focused attention on a specific criterion of directing technique. Using those questions, they were then asked to evaluate the production on five-point equal-interval scales.

An important consideration must be inserted here: any reliability coefficient is as much a reflection of the test's ability to elicit consistent responses as it is of the judges' abilities to maintain consistent patterns of judgment. Consequently, the attempt in these questions was not to experiment with some projective technique to improve test reliability; but, to elicit as succinctly as possible judges' direct responses to the function of each of the 15 concepts in performance.

The following were the directing criteria used in the study:

Unity	Characterizations	Business
Variety	Ensemble	Emotion/Key
Meaning	Relationships	Mood
Style	Composition	Tempo
Story line	Movement	Rhythm

Finally, the judges were asked to rank the quality of the three performances. No ties were allowed in the rankings.

Results

The raw data consisted of 60 "overall evaluations," 900 ratings of "specific criteria" (that is, ten judges rating 15 criteria for six performances), and 60 rankings.

Procedures of analysis used in this study were similar to those employed by Richardson and Waal. Because of the preliminary nature of the study, the .05 level of confidence was set as the limit for statistical significance.

Hypothesis 1:

Hypothesis 1 stated that judges rate "specific criteria" of directing excellence more reliably than they rate "overall performance." The implication was that error variance could be reduced by judging the varying aspects of a directing performance independently. Richardson and Waal, though they did not prove it in their own study, indicated, by comparison with the Clevenger and Tolch experiments, that such was the case in judging actors' performances.

Coefficients of reliability were computed using Ebel's analysis of variance technique for estimating reliability by intraclass correlation (Guilford, 1954, pp. 395-397). The coefficients are displayed in Table 1. Two reliability

TABLE I

RELIABILITY ESTIMATES OF RATING SCORES

SPECIFIC CRITERIA	TRIAL I		TRIAL II	
	r_1	r_k	r_1	r_k
Unity	.79	.97	.34	.84
Variety	.52	.91	.38	.86
Meaning	.60	.94	.73	.96
Style	.66	.95	.28	.80
Story line	.80	.98	.73	.96
Character	.48	.90	.49	.90
Ensemble	.75	.97	.21	.73
Relationships	.70	.96	.19	.70
Composition	.71	.96	.00	.00
Movement	.76	.97	.44	.97
Business	.80	.97	.10	.53
Emotion/Key	.61	.94	.25	.77
Mood	.73	.96	.15	.64
Tempo	.80	.98	.22	.74
Rhythm	.79	.97	.34	.84

coefficients for each trial (each bill of one-acts) were given: the first coefficient (r_1) being the index of reliability for any given single judge; the second coefficient (r_k) being the reliability of the mean of all (k) ratings. The fact that r_k was consistently greater than r_1 offered little more than corroboration of the forgone conclusion that "most gains in reliability come from multiplying raters when initial reliability is low . . ." (Guilford, 1954, p. 397). The amount of difference between r_1 and r_k should be the subject of considerable concern to even the most well-qualified directors and teachers, particularly when academic success or failure rests on the accuracy of singular qualitative evaluations.

In comparing Trial I with Trial II, results similar to those of the Richardson-Waal study were apparent. One of the bills of one-acts (the second) was much more evenly matched in terms of excellence. Consequently it was more difficult for judges to evaluate it consistently. As was illustrated in the earlier study, the ability of judges to discriminate seems to be related directly to the amount of variance between performances, regardless of whether the evaluations are of acting or of directing.

To complete the test of the first hypothesis, however, it was necessary to compute reliability coefficients for "overall evaluations." The results of those tests are presented in Table 2.

When the results of those tests were compared with the reliability estimates from Table I, the first hypothesis was rejected. Although no tests were available to determine significance levels, the similarity of the estimates in-

TABLE 2

RELIABILITY ESTIMATES OF "OVERALL EVALUATIONS"

	r_1	r_k
TRIAL I	.81	.98
TRIAL II	.15	.64

dicated neither an advantage nor a disadvantage in specifying directing criteria as a means for increasing reliability. One might even suggest that the estimates for Trial I of the "overall evaluations" were equal to or higher than their counterparts in Table 1.

Rejection of this hypothesis did not confirm the findings of Richardson and Waal. A variety of reasons might be suggested. One possible explanation was the difference in the way in which the sets of questions were phrased in the two studies. A second, and more obvious explanation may be found in the difference between the phenomena being evaluated. Although the broad framework of judgment was the same, the object of the judges' concentration was different. Whether or not directing is an area which lends itself more suitably to generalized evaluation than does acting can only be conjectured, and should be investigated in further comparative research. A final suggestion, however, lies within the Richardson-Waal study itself. Their conclusion was based upon a comparison of reliability coefficients from two different populations--their own and the Clevenger-Tolch group. This may well have distorted their inference. "One should speak of the reliability of a certain instrument applied to a certain population under certain conditions." (Guilford, 1965, p. 439)

Hypothesis 1 was not confirmed. There was little if any difference in the reliability of judges' ratings of "specific criteria" and "overall performance."

Hypothesis 2:

Hypothesis 2 stated that judges rank directing performances more reliably than they rate either "specific criteria" or an "overall performance." Kendall's coefficient of concordance (Siegel, 1956, pp. 229-238), yielding a reliability estimate comparable to Ebel's intraclass correlation, was used to establish the reliability of the rankings--the final response item on the test. The rankings illustrated even more clearly the problem of judging evenly matched performances. Complete unanimity of rankings on Trial I produced an unusual $W = 1.00$; while Trial II yielded a $W = .520$, the latter was not statistically significant.

These statistics, however, had to be put on a comparative basis with two other sets of reliability estimates in order to resolve the question raised in the hypothesis. First, in relation to the reliability of "overall performance," ranking compared favorably: the "ranking" estimates of reliability; for Trials I and II were $W = 1.00$ and $W = .520$ respectively; whereas, the "overall performance" estimates were $r_1 = .81$ and $r_1 = .37$. It should be noted that the Kendall

coefficient is comparable to r_1 rather than to r_k since it indicates the percentage of true variance shared commonly by each individual judge.

Secondly, in comparing the reliability estimates of the "rankings" with the "specific criteria" coefficients, further evidence was offered favoring acceptance of the second hypothesis. R_1 for Trial I, ranged from $r_1 = .48$ to $r_1 = .80$, compared with "rankings" $W = 1.00$. In Trial II only two out of the 15 "specific criteria" coefficients--both with $r_1 = .73$--ranged above $W = .520$. The evidence suggested a cautious acceptance of Hypothesis 2, supported somewhat by a proportionate set of estimates in the Richardson-Waal study. As with that earlier study, however, the differences could have occurred by chance as no statistical procedures exist to assess their significance.

Hypothesis 3:

Hypothesis 3 stated that night-to-night shifts of intuitive standards of judgment will be affected by specifying criteria.

Two analyses of variance were required to test this hypothesis: the first involved the judges' evaluation of the "overall performance"; and the second, the judges' evaluation by "specific criteria." In both cases a two-factor design was used employing repeated measures on judges (Winer, 1962, pp. 302-312). The "specific criteria" ratings for each judge were summed to form the criterion measure in the second variance analysis. The results of both analyses may be found in Tables 3 and 4.

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF EVALUATIONS OF "OVERALL PERFORMANCE"

Source of Variation	Sums of Squares	df	Mean Squares	F
Between Subjects		5		
A (trials)	.62	1	.62	.07
Subjects within groups	33.73		8.43	
Within Groups		54		
B (judges)	1.42	9	.16	.50
AB (judges by sessions)	3.38	9	.38	1.19
B x Subjects within groups	11.60	36	.32	

*Significant at .05 level of confidence.

The analyses suggested the following conclusions relative to the third hypothesis: The judges-by-trials interaction effect in Table 3 was not significant suggesting that the variances in the mean ratings were not affected by night-to-night shifts of intuitive standards of "overall evaluations." It is apparent in Table 4 that the same was true also of the variance of "specific criteria" judgments. Consequently, Hypothesis 3 was rejected. There appeared to be no affect on night-to-night judgment variance as a result of specifying directing criteria.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF "SPECIFIC CRITERIA"

Source of Variation	Sums of Squares	df	Mean Squares	F
Between Subjects		5		
A (trials)	123.26	1	123.26	.07
Subjects within groups	6669.67	4	1667.42	
Within Groups		54		
B (judges)	287.66	9	31.96	.75
AB (judges by sessions)	196.41	9	21.82	.51
B x Subjects within groups	1528.33	36	42.45	

*Significant at .05 level of confidence.

Again based on a comparison with significant interaction found in the Clevenger-Tolch study, Richardson and Waal concluded, by virtue of their non-significant F, that night-to-night judgmental inconsistencies could be reduced by specifying criteria. The results presented here tended to be consistent with neither of the two earlier acting studies: evaluations of "overall performance" of judges did not tend to be affected by night-to-night shifts, nor was the variance affected significantly by the addition of "specific criteria."

Discussion

The following conclusions may be drawn from the findings of this study: (1) judges of directing appear to be as reliable in evaluating "overall performance" as they are when they rate "specific criteria," (2) judges probably rank "overall performance" more reliably than they can rate either "specific criteria" or "overall performance," and (3) specifying directing criteria has no effect on night-to-night shifts of intuitive standards of judgment.

In almost all cases where more than a single judge was involved, reliability coefficients were higher than might have been expected; in only a few instances did error variance account for most of the variability in the scores. This seems to suggest that a group of competent judges can evaluate the work of directors of varying ability with a relatively high degree of reliability.

This study was concerned with the reliability of judgments of directing. No attempt was made to measure validity; however, increases in reliability usually result in subsequent increases in validity.

The coefficients of reliability revealed in this study were unusually reliable when compared to similar coefficients found in studies of the reliability of ratings of acting. One of the most striking results was that there was very little difference between the reliability of "overall evaluations" and judgments of "specific criteria." Although these findings seem to favor the simpler method of evaluation, we feel the use of "specific criteria" has value for instructional purposes, and for identification of areas of inconsistency in judgments.

REFERENCES

Dr. Kepke is Assistant Dean of Arts and Sciences at Bowling Green State University. Dr. Stockwell is Assistant Professor of Drama at San Fernando Valley State College, and Dr. Zyromski is Assistant Professor of Drama at Virginia Polytechnic Institute. This article is based on a paper read at the annual convention of the American Educational Theatre Association, Los Angeles, August 1968.

Clevenger, T., Jr., Tolch, J. C. The Reliability of Judgments of Acting Performance. Educational Theatre Journal, 1962, 14, 318-323.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1954.

Richardson, D., Waal, C. Increasing the Reliability of Judgments of Acting Performance. Quarterly Journal of Speech, 1966, 52, 378-382.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

ACTOR-CHARACTER PERSONALITY IDENTIFICATION IN A THEATRE PRODUCTION

R. Wayne Smith

The degree to which an actor does or does not identify with the roles he assumes for the stage is a subject of serious historical and theoretical importance. Although Stanislavski's appealing campaign favoring the actor who "lives the part" (1945) has given the controversy a peculiarly contemporary flavor and relevance, the argument is by no means the exclusive property of modern theory or theorists. To the contrary, it claims a long and healthy history reaching all the way back to the very roots of the art of acting itself. If the portrayal is sufficiently skilled to convince us, the audience, of the "truth of the characterization," then it is only natural for us to wonder what affects the portrayal must have on the actor's self or psyche. And for that reason, people often speculate regarding the schizophrenia of the actor of many parts.

For all its popularity and theoretical significance, the topic has never been explored more than tentatively in an objective manner or setting. Within a study focusing primarily upon the "semantic compatability of actors and director," Allen N. Kepke (1963) used Q-Methodology to investigate actor-character identification. His findings suggested that there was a limited tendency for some actors to perceive their stage roles in terms of their own perceptions of themselves. That exploration, however, was not the primary concern of the study, hence only limited attention was given to the quantity or quality of change in perception of self and character progressively throughout rehearsals and performance.

The present study was an attempt to expand our understanding of the dynamics of actor-character personality identification using a larger sample and by obtaining more repeated measures during rehearsals and performance. Specifically the study was an attempt to detect and explain similarities in actors' perceptions of themselves (self-concepts) and their perceptions of their stage roles or characters (character-concepts).

METHOD

Both perceptions (self-concept and character-concept) were measured similarly using Raymond G. Smith's (1964) semantic differential for theatre concepts. Although that differential has been the target of considerable criticism (Clevenger, et. al., 1967, 241-246; Hansen, 1966; and Thayer, 1964, 180-83), it was considered appropriate to the present study inasmuch as the study was concerned more with the differences between subsequent measures obtained from the same scales than it was with the interrelationships among the scales, factors, and concepts. The following four Factors are represented by the following 11 scales in Smith's differential: Factor I (Manner): calm-excitable, cold-hot; Factor II (Seriousness): light-heavy, serious-humorous, tense-relaxed; Factor III (Ethical Value): honest-dishonest, worthless-valuable, true-false; Factor IV (Esthetical Value): ugly-beautiful, displeasing-pleasing, painful-pleasurable. Operationally, any two concepts were considered compatable whenever the distance between the majority of the scales defining the concepts was not statistically significant at the .05 level of confidence. Identification was assumed whenever one of the two compatable concepts was an actor's self-concept and the other was that same actor's character-concept.

The experimental subjects were ten undergraduate and two graduate students at Bowling Green State University. All were members of the cast of The Garbage Hustlers, a three-act comedy by George Savage, Sr. and George Savage, Jr., a Summer Theatre Production at Bowling Green State University, 1969. Since the Bowling Green performance of The Garbage Hustlers was the first in Ohio, it was assumed that none of the subjects was familiar with the play or any of the characters before they were given a script by the experimenter. Try-out scripts were given to aspirant cast members only after they had completed the first differential; the first self-concept measure.

In all, nine measurements were made: five self-concept measures, and four character-concept measures. To obscure the intent of the study, character-concept measures were intermixed randomly with similar forms requesting descriptive concepts of other characters in the play. Scores obtained from the latter concepts were not tallied. To obscure the intent of the study further, the measurement of the final character-concept preceded the measurement of the final self-concept by 24 hours. The concepts and the times at which they were measured are displayed below:

Measure	Description of Measure	Type of Rehearsal	Date of Measure
A_1	Self-concept	Before Tryout	6/23, 24, 25/69
A_2	Self-concept	1st Read-thru	6/26/69
C_1	Character-concept	1st Read-thru	6/26/69
A_3	Self-concept	1st Run-thru	7/8/69
C_2	Character-concept	1st Run-thru	7/8/69
A_4	Self-concept	Final Dress	7/16/69
C_3	Character-concept	Final Dress	7/16/69
C_4	Character-concept	2nd Performance	7/18/69
A_5	Self-concept	3rd Performance	7/19/69

Using the formula, $D = \sqrt{\sum d^2}$ (where d is the difference between the average of the scales comprising the factors of any concepts compared), semantic distance scores, or D-scores, were computed in order to compare the actor's original self-concept measurement with all other self-concept measures. The results of those comparisons are displayed in Table I.

Table I lists Actors in Column 1 and D-score comparisons in Columns 2 through 5. For example, a D-score of .81 is listed for Actor 1 for comparisons A_1/A_2, indicating that Actor 1's initial self-concept measure obtained just before tryouts (A_1) differs from his self-concept measure obtained immediately following the first read-through of the play (A_2) by an average scale value of .81. According to Smith (1961, p. 6), D-score and Factor shifts ≥ 1.5, and individual scale shifts ≥ 2.0 may be considered to be statistically signif-

TABLE 1

D SCORES: ACTOR'S SELF-CONCEPT COMPARISONS

Actor	A_1/A_2	A_1/A_3	A_1/A_4	A_1/A_5
1	0.81	1.00	0.90	1.50*
2	0.90	1.99*	1.21	0.69
3	0.95	0.69	0.76	0.58
4	0.68	1.12	0.75	1.25
5	0.96	1.25	0.33	0.50
6	1.96*	1.99*	1.97*	1.06
7	0.47	0.90	0.60	1.00
8	0.60	0.80	1.05	1.00
9	3.90*	4.41*	4.15*	3.68*
10	1.45	3.31*	3.40*	5.30*
11	3.32*	1.42	1.37	2.35*
12	2.08*	2.64*	2.77*	5.00*
Mean	1.51*	1.79*	1.61*	2.02*
Standard Deviation	1.11	1.15	1.22	1.71

*Significant semantic distance at the .05 level.

icant at the .05 level of confidence. For Actor 1, comparisons A_1/A_2, A_1/A_3, and A_1/A_4 do not differ significantly, while comparison A_1/A_5 may be considered as a comparison of two statistically different concepts. The self-concept of Actor 1 remained stable through the first three measures and changed significantly when measured the fifth time.

The Table indicates that in 30 of the 48 comparisons made, the distance between measures was not statistically significant. Apparently self-concept for this group was a reasonably stable measure. It is also apparent that when significant shifts were made they usually occurred with the second measure and thereafter remain relatively stable. That pattern was particularly evident in the shifts recorded for Actors 6, 9, 10, and 12.

Similarly the initial character-concept measure was compared with all other character-concept measures, and the results of those comparisons may be found in Table 2.

The differences between the character-concept measurements revealed in Table 2 were generally wide and variable; 28 comparisons were found to be statistically significant. When those changes are compared with the changes in self-concept (Table 1), it is apparent that actors' perceptions of their roles or characters are a good deal more variable than are perceptions of self.

D-scores were also computed to compare the actors' initial self-concepts (A_1) with successive measures of character-concepts (C_1, C_2, C_3, and C_4).

TABLE 2

D SCORES: COMPARISONS OF CHARACTER-CONCEPTS

Actor	C_1/C_2	C_1/C_3	C_1/C_4
1	1.05	1.12	1.21
2	1.42	0.94	1.97*
3	2.78*	2.00*	3.04*
4	2.56*	4.67*	3.54*
5	0.58	0.96	2.19*
6	1.50*	1.50*	1.50*
7	5.05*	3.54*	3.54*
8	2.08*	1.12	1.68*
9	2.67*	2.83*	4.62*
10	3.35*	1.77*	2.67*
11	2.56*	2.78*	1.53*
12	2.45*	2.58*	2.17*
Mean	2.34*	2.15*	2.47*
Standard Deviation	1.17	1.16	1.03

*Significant semantic distance at the .05 level.

Those comparisons are displayed in Table 3. The first comparisons reported in Table 3, all of which were significant, indicate that all of the actors were assuming roles which they considered different from their self-concept. When we observe successive measurements no consistent pattern emerges. The discrepancy between the Self and the Character widens in some cases and decreases in others, and at the time of the final measures, nine of the 12 comparisons still remain significantly different. Of the seven significant shifts ($\geq$ 1.50) from one measure to another, five were changes associated with measures taken at a dress rehearsal. Other researchers have reported similar concept shifts when concepts were measured at that time (Hansen, 1966; Kepke, 1963; and Tannenbaum et. al., 1963).

Finally, D-scores were computed for concurrent measures of self-concept and character-concept. The results of those comparisons are to be found in Table 4.

Although Table 4 records fewer changes than Table 3, the changes appear to be more concentrated and more consistent. Once again it is apparent that the greatest number of changes are associated with measures taken before or after those made at the dress rehearsal.

Each set of comparisons made within each of the following Tables was subjected to the Wilcoxon matched-pairs signed-ranks test for correlated samples to assess the significance of change (Ferguson, 1966, pp. 360-61). Those tests revealed that none of the changes compared in Tables 1 (Actor's self-concept com-

TABLE 3

D SCORES: ACTOR'S SELF-CONCEPT COMPARED WITH CHARACTER-CONCEPT

Actor	A_1/C_1		A_1/C_2		A_1/C_3		A_1/C_4
1	1.54*		0.68		0.82		1.49
2	2.17*		2.80*		2.37*		1.25
3	2.18*		1.73*		0.90		0.96
4	4.48*	→	2.64*	→	0.96	←	2.58*
5	1.60*		1.70*		1.07	←	3.51*
6	5.52*		5.31*		5.31*		5.31*
7	3.68*	→	1.76*		2.47*		2.47*
8	2.52*		3.95*		2.87*		3.73*
9	1.95*		3.31*		3.81*	←	5.87*
10	4.44*		2.17*		3.63*	←	5.39*
11	3.51*		2.41*		2.00*		2.42*
12	1.97*		2.19*		1.07		2.16*
Mean	2.96*		2.55*		2.27*		3.10*
Standard Deviation	1.32		1.21		1.43		1.68

*Significant semantic distance at the .05 level.
←Shift of significant magnitude at the .05 level of confidence.

TABLE 4

D SCORES: COMPARISON OF CONCURRENT MEASUREMENTS ACTOR'S SELF-CONCEPT AND CHARACTER-CONCEPT

Actor	A_2/C_1	A_3/C_2	A_4/C_3	A_5/C_4
1	1.61*	1.21	1.50*	1.54*
2	2.55*	1.38	3.47*	1.68*
3	1.46	2.27*	1.10	1.38
4	2.69*	2.52*	1.21	1.50*
5	2.00*	2.26*	0.90	3.32*
6	4.99*	5.51*	4.81*	5.52*
7	2.24*	1.77*	2.19*	3.38*
8	3.93*	2.68*	2.35*	4.23*
9	2.62*	3.36*	4.87*	2.11*
10	2.93*	2.98*	3.57*	1.17
11	1.77*	1.05	0.96	1.29
12	1.51*	2.02*	2.35*	2.67*
Mean	2.53*	2.42*	2.38*	2.48*
Standard Deviation	1.05	1.20	1.40	1.38

*Significant semantic distance at the .05 level.

parisons), 2 (Comparison of Character-concepts) and 4 (Comparison of Concurrent Measurements: Self-concepts and Character-concepts) were statistically significant.

In Table 3, however, where the initial self-concept was compared with successive measures of the character-concept, two sets of comparisons were found to differ. Comparisons of the initial self-concept with the third and fourth character-concept measures (A_1/C_3, and A_1/C_4) were found to differ significantly. When comparisons were made with the final performance measures, the differences were not statistically significant.

We must conclude, then, that generally the actors used in this study did not identify with their stage roles to any appreciable or detectable extent.

Furthermore, no evidence was revealed to suggest general and progressive identification from tryouts through public performance. Though the analyses did suggest progressive development of identification from tryouts through the final dress rehearsal, that progress was offset by regressive development from dress rehearsal through the third performance.

Although the findings are clear with regard to the quantity of the Actors' identification with his character, they say very little about the actors' patterns of identification with their roles. In an attempt to provide more detailed information regarding patterns of association, each actor's initial and final self-concept and initial and final character-concept were plotted on each of the 11 scales of the semantic differential. An illustration for Actor 10 is plotted below:

0 Actor's original concept.
● Actor's final concept.
▯ Character's original concept.
▮ Character's final concept.

The four concepts plotted on 11 scales yielded 528 concept positions on 12 semantic differentials. The analysis led to the identification and measurement of 132 separate movements; 12 actors on 11 scales. When the concept movements were counted, it was found that 71 changes in the actor concepts were significant (scale moves must be $\geq$ 2.0 to be considered significant), while 61 were not significant. The percentage (54.8%) of significant moves was not in itself significant.

Further, a descriptive analysis and count of the direction of moves was conducted. The frequencies and directions of significant moves are displayed in Table 5. A chi-square test comparing the observed and expected frequencies of

TABLE 5

CHARACTERISTIC CONCEPT MOVEMENT

Type of Movement	Frequency
Self-concept away from Character-concept	8
Character-concept away from Self-concept	5
Character-concept and Self-concept Toward Each Other	3
Character-concept toward Self-concept	33
Self-concept toward Character-concept	22
Total Moves	71

moves resulted in the rejection of the hypothesis that there were no significant differences in the five patterns of movement identified in Table 5 (Ferguson, 1966, pp. 192-93). Two significant patterns were clearly identified: Either the actor's perception of himself moved in the direction of his character, or the character moved in the direction of the self. That is to say that when movements do occur, they are generally toward either "identification with the role" (movement of self-concept toward character-concept), or toward what Stanislavski (1945) called "exploitation of the role" where the actor molds the role to suit his own personality (character-concept moves toward self-concept). Whatever the case, moves generally take the direction of resolving the semantic gulf between the self and character; other types of moves are the exception rather than the rule.

The above discussion clearly reflects conflicting results. For the group, semantic compatability was never achieved, yet significant patterns of movement were detected and identified. The similarity of information revealed in the concept plots of the individual actors and in the group comparisons of D-scores suggests that discrepancies in the way in which actors develop their roles are readily discernable and perhaps worthy of further study. Because of the discrepancies in the talent and experience typical in student casts, it is possible that more homogeneity of approach might be evident in more experienced casts. At present that is clearly a subject suitable only for speculation. Inasmuch as the cast studied in this investigation was not unique, it is likely that the discrepancies in approach to role reflect student actors' very real and intense struggles to master their stage roles.

The following conclusions are tentative: (1) The way in which an actor perceives his relationship with his role in a play is amenable to objective analysis. (2) During the process of rehearsing and performing a play, changes in an actor's self-concept are only minor and are not statistically significant. (3) As the actor's role takes shape it does undergo significant changes and modifications. (4) There is no clear evidence to suggest that actors ultimately identify with the roles they play on the stage. (5) Although actors in general do not identify with their stage roles, some individual actors do. (6) Two patterns of movement are clearly evident. Both tend to reduce the semantic space between the actor's concept of himself and that of his character: the character moves toward the actor's self-concept or the self moves toward the character.

REFERENCES

Dr. R. Wayne Smith is Assistant Professor of Drama at Fresno State College, Fresno, California. Originally this study was a class project under the direction of Professor David W. Addington, Bowling Green State University. The author wishes to gratefully acknowledge the co-operation of the director and cast members of The Garbage Hustlers.

Clevenger, T., Jr., Clark, M. L., & Lazier, G. N. Stability of Factor Structure in Smith's Semantic Differential for Theatre Concepts. Quarterly Journal of Speech, 1967, 53, 241-246.

Ferguson, G. A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill, 1966.

Hansen, B. K. The Semantic Differential as a Measure of Meaning in the Theatre. (Doctoral dissertation, University of Minnesota) Minneapolis, Minn.: University Microfilms, 1966, Number 67-7732.

Kepke, A. N. A Study of Communication of Perception of Character Among Actors, Directors, and Audience Using Q Methodology. (Doctoral dissertation, Michigan State University) Lansing, Mich.: University Microfilms, 1963, Number 64-7520.

Rabby, L. B. An Application of Smith's Semantic Differential for Theatre Concepts to Character Judgments. Paper presented at a meeting of the American Educational Theatre Association, Los Angeles, August 1968.

Smith, R. G. A Semantic Differential for Theatre Concepts. Speech Monographs, 1961, 28, 1-8.

Stafford, H. F. Theatre Practice: The Art of Acting. Theatre Arts. 1934, 18, 527-532.

Stanislavski, C. An Actor Prepares trans. by Elizabeth Reynolds Hapgood, New York: Theatre Arts, 1945.

Tannenbaum, P. H., Greenberg, S. S., & Leitner, M. A. Changes in Semantic Compatibility During the Production of a Play. Speech Monographs, 1963, 30, 340-344.

Thayer, D. Use of Smith's Semantic Differential for Theatrical Performance Concepts. Speech Monographs, 1964, 31, 180-183.

PSYCHOLOGY AND DRAMA

Brian Sutton-Smith and Gil Lazier

During the Spring of 1969, the present investigators, professors of psychology and theatre respectively, initiated a joint exploration of some of the antinomies between the social sciences and art. We were interested particularly in the parallels between the "languages" of psychology and drama. The work is illustrated by some preliminary methodological attempts to develop criteria for rating one particular parallel; the concept of "involvement."

Several pairs of terms may serve to illustrate the potential fruitfulness of the collaboration we hope to illustrate.

Rehearsal or Socialization: Actors use many methods for "getting into a role." Psychologists talk about the way in which children get into their roles as they grow older, but neither seems to take much notice of the other. For example:

(a) The actor may be helped into a role by reading about it or by hearing others discuss it. Though commonplace in the theatre, the method has never been popular with psychologists who generally rely upon less audible means for conveying information. Nonetheless it is still a method popular among parents attempting to socialize adolescents, and it may also be popular among adolescents attempting to socialize themselves.

(b) Although forms of imitation or modelling are popular methods among psychologists for helping children learn their roles, the methods are usually abhorred by theatre people who consider them the methods of "dictatorial directing," and denigrate them by calling them "giving line readings," or even "Mickey Mousing."

(c) Actors are often rewarded and punished by directors in much the same way as children are rewarded and punished by their parents, but those techniques are more openly acknowledged in socialization theory than they are in theatrical practice.

(d) Directors sometimes stimulate actors to experience an emotion by using some of the following strategy: they make the actor experience terror by frightening him, exhaustion by having him run around the block, anger by slapping him, etc. These techniques, though relatively common to the stage, are foreign to the psychologist in general practice.

(e) Another technique common to the actor is the evocation of the actor's feelings and actions during a prior and parallel experience in order to use those feelings and actions according to the demands of a current role; sometimes called "sense memory," "emotion memory," or simply "recall." It is probable that this is a primary tactic used by children to accomodate novel requirements by assimilating them to some prior experience.

While we have illustrated only a few of the parallels between the rehearsal process and the socialization process, the fruitfullness of reconsidering each in terms of the other should be apparent. While psychologists usually consider each technique as distinct from all others, the theatre director tends

to employ as many rehearsal techniques as possible in an attempt to achieve his ends. It is possible that parents proceed in the same way.

Getting Into Character or Ontongenesis: Not unrelated with the above discourse are changes which occur in actors through extended rehearsal, and in children throughout the course of their development toward adulthood. Whether the development of the actor in a role parallels the growth of a child in development is unknown, but the similarities are certainly worthy of further study. We might ask, for example, is the basic line of development from the diffuse to the particular, or from caricature to authentic representation?

It seems probable that actors like children first perceive salient gestalts and then attempt to reproduce them. If that assumption is correct, the methods of teaching creative dramatics may be faulty, as they tend to emphasize analysis as a means for understanding roles, rather than more generalized or original reactions to the totality. It could be that proponents of creative dramatics falsely proceed from the part to the whole rather than the reverse. The child's excited, if caricatured, representations may be the more proper starting point.

Some evidence to suggest this has been generated by Neil Martin who has been experimenting with a variety of procedures in dramatic improvisations specifically with black and Puerto Rican functional illiterates. He began by using the standard creative dramatics practice of small, partial, isolated exercises, culminating finally in a full story improvisation (Berger, 1966; Siks, 1958; Ward, 1957) but found this process totally unsuccessful. That failure led him to attempt a more unorthodox approach; presenting a total dramatic story to his participants encouraging them to "do it all at once." His success was immediate (Lazier, 1969).

Projection vs Projection: As a stage technique, projection is a term used to describe the size or force of the actor's affect in a given environment. When a performer is criticized for "overprojection," we may assume that his visual and/or aural message is too large for the size of the playhouse; he is expending too much energy for the space. Further, each actor has his own relative level of involuntary projection. Some actors seem to "radiate" more, or have a larger "sphere of influence" (Stanislavski, 1924, 1936, 1949) than others.

In psychology, projection is used to describe the act of attributing to others or to objects one's own feelings, desires, or motives. By telling stories about pictures, playing with dolls, or by painting pictures a subject is encouraged to believe that he is permitted to disguise his "true" self, but without realizing it, he actually betrays his underlying affects by "projecting" them into stories austensibly about someone else.

At the very least dramatists and psychologists should clarify the way in which they use the same terms. One meaning clearly emphasizes overt and desired capacities while the other denotes covert and involuntary revelations. The psychologist might, nonetheless, contend that within the actor's projections may be found expressions of his own psychological character, so that the one form of projection may in fact contain the other.

The Assessment of Dramatic Involvement

Like projection, involvement is a term used by dramatic theorists and psychologists alike with entirely different intent. In some pilot video work, for example, we discovered that our definitions varied considerably.

The psychologists were rating children as "involved" when they had the appearance of what Goffman (1963) might term an "away" state. Following Sarbin (1954), psychologists might rate degrees of involvement along a dimension proceeding from the highest point to the lowest: from voodoo death, through rites of passage (ecstasy), hysteria, hypnosis, heated acting, mechanical portrayal, to customary roles.

Theatrical practitioners, on the other hand, work with more specifically theatrical concepts. The actor is "involved" not only when he appears "lost in the role," but also when his specific actions fulfill the requisites of dramatic portrayal.

The precise statement of the character of that involvement is the major focus of this study. We would argue that if involvement in drama could be reliably assessed, then this dependent variable could be used to tell us many other important things about participation in the drama.

The vehicle for study was naturalistic observation of a normal creative drama program with three groups of children ages eight, nine, and ten, with approximately ten subjects in each group. The structure and content of experiences followed the lines indicated in standard creative dramatics texts (Berger, 1966; Siks, 1958; Ward, 1957). While this program was in progress two observers kept detailed records of the behavior of the instructor including such things as verbal content, gestures, mimes, and evoked imagery and memory, and paying special attention to the full variety of role induction procedures as they were employed with specific dramatic events.

Other observers regularly watched the children and simultaneously developed the rating scales for dramatic involvement described in the following pages.

Graduate students in the theatre were used for the assessment and recording of the role induction procedures and psychology students were used to assess role involvement. The two recording procedures remained relatively independent with as little communication between the teams as was needed for calibrating such things as times, dates, and episodes, and to permit subsequent generalizations about the relationships between the two sets of data: role induction and role involvement.

Certain phenomena were immediately apparent. It was clear that role involvement was highly susceptible to the effects of role induction, that some procedures were more effective than others, and that the personality of the participant played a rather crucial interactive role in the process. At this stage of the research, however, it seemed important to concentrate on the assessment of involvement saving personality interaction for investigation at another time.

The following scales are offered only as a tentative model for a

measuring instrument, and were the product of many hours of argument and discussion. Though seven intervals were used to separate the scale dimensions, the numerical choice was more a matter of arbitrary desire than of methodological design.

1. Focussed-Distracted

Initially, the focussed-distracted continuum was one of our most useful scales, because the primary problem in the early sessions seemed to be focus of attention. The measure indicated global focus of attention including general task involvement as well as completion of the activities. Later, as more of the children became capable of staying in field during acting, the scale was revised to be more specifically concerned with how engrossed the child was in the dramatic task; how convincingly he seemed to be engaged in the performances of his "as if" behaviors. The child rated "7" would reveal a constantly inward focussing of attention, would never look at others for cues or attention, would seem to be totally engrossed in what he was doing. The child rated "6" would behave similarly, but with somewhat less intensity and perhaps a slightly delayed warm up. A "4" would denote a median performance and a comparatively greater degree of variability in concentration and perhaps a measure of "hamminess" which might have constricted the focus. The child rated as "1" might look to the audience or to other actors for cues, might not create a convincing world of imagination, and might be distracted easily by external cues or irrelevant internal stimuli.

2. Completes-Truncates

This scale was used to describe the degree to which the child completed the basic task. Did he include the major points of the dramatic problem; did he develop each piece of business thoroughly; was his pacing sufficiently relaxed to allow the completion of each part, or did he appear to want to quit the acting arena early? This scale differed from another, "elaboration," to the extent that the business need not have been innovative as long as it was developed adequately enough to be easily understood. A rating of "7" indicated a complete presentation which included all the necessary dramatic details and pacing appropriate enough to allow for the adequate development of subtle detail. A rating of "4" might be used to indicate the omission of some parts of the task, but inclusion of the major elements, inappropriately rapid pacing, or the skimming of detail. It might also be used to indicate a performance which truncated the beginning or ending of the work. A rating of "1" was reserved for performances which revealed clear omissions of a serious nature.

3. Use of Imaginary Objects: Consistent-Inconsistent

This scale was particularly appropriate to pantomime. It was meant to be used as a measure of the capacity for creating convincing and palpable objects through physical means and conveying their properties, such as size, texture, weight, temperature, function, and shape in a consistent manner. Factors determining a high rating would include the number of relevant properties created and maintained consistently as well as the appropriateness of the gestures used to communicate credibility (no exaggerated or stereotyped movements). In young children, objects which are heavy are often lightly lifted, doors which are shut are walked through a moment later. The appropriate actions for the created object and respect for its character were important here.

Our problem with this variable was to determine whether we were dealing

with the consistency or with the presence of this behavior. Although some children were consistent in the way in which they used imaginary objects, the number of objects they created was limited. Hence another category was needed to describe the number of imaginary objects created. The next category served that purpose.

4. Elaboration-Lack of Elaboration

This scale included the elaborative creation of new ideas in addition to those requisite in the description of the dramatic task. It measured innovative approaches to solving the problems presented by the task. It was used also to describe the communication of additional or unusual facts about the imaginary objects used; detail which might render the objects more singularly vivid to the viewer. Consideration was also given to the relevancy of the elaboration or detail, and while "irrelevant elaboration" might be useful in evaluating the performances of experienced actors and while it is clearly the opposite of "relevant elaboration," the usefulness of that description was found to be questionable when applied to the process of evaluating the performances of children. Even though excellence and relevance were joined in the present work, they may warrant separation in further research.

5. Use of Space: Appropriate-Poor

On this scale, high scorers used enough space to include all objects needed to allow for free bodily movement and to provide for variety, but not so much space as to become vague or undefined. Low ratings reflected the use of an unnaturally constricted area, poorly defined space, and a tendency to use only one part of a larger area even though more space might have been available.

6. Facial Expressions: Appropriate-Inappropriate

This scale was used to rate the ability of the child to convey the emotions of a character in a convincing and credible manner within the limits of the dramatic task. Reactions should be relevant, appropriate in intensity (neither too restrained nor too exaggerated or stereotyped), and show enough variety to maintain interest and to suggest consistent involvement. Low ratings were used to indicate facial expressions which reveaied inadequate variety or intensity, exaggeration, stereotyped response, or inappropriate response.

7. Body Movements: Appropriate-Inappropriate

This scale was used to measure the effectiveness of bodily movement in terms of its relevance to the given dramatic situation, its appropriateness for a particular character (age, status, feelings, etc.), and its success in communicating contextual messages. Constricted movements (unless appropriate to the character), or flaccid posture tend to diminish the child's stage presence and his ability to hold viewer interest. A high rating on this scale reflected the child's freedom of movement in space, his variety, velocity and contour of movement (angular vs curved), and the relevance of this behavior to the dramatic task.

8. Vocal Expressions: Appropriate-Inappropriate

This scale measured the effectiveness of verbal and vocal expression. Like facial expressions, the voice conveys the emotional reactions of the character to the audience. It also reflects the child's ability to improvise elaborate and spontaneous verbal responses within the dramatic task. A high score

indicated clear vocal projection, variety, and emotional relevance.

9. Social Relationships: Appropriate-Inappropriate

This final scale was intended to measure the actor's awareness of and reaction to other participants in the dramatic situation. To receive a high rating the child was expected to involve himself with others in a convincing manner, to react to others, to pace himself appropriately with others, and to reveal his consciousness and awareness of the others in the group. Typically the interactions of those who scored poorly were inappropriately timed, repetitive, or inadequate. Those who scored poorly did not appear to be able to properly judge or respond to the needs and desires of others.

Some Psychological Considerations

It is clear from the above discussion that our variables remain inadequately refined. Nonetheless we found them reasonably appropriate for use with this group of young children involved chiefly in pantomime and improvisation. Whether they would be equally appropriate for use with adults or in other acting situations is questionable.

The preliminary analysis of the ratings indicates an acceptable degree of reliability across all observers. But the quantitative data will be presented in a subsequent paper in which the relationship between involvement and the type of dramatic task will be described and developed in more detail.

Turning again to the scales, we are impressed with the great variety of psychological studies they suggest. For example, it seems likely that children vary greatly in their ability to focus in drama. Distractibility, impulsivity, and hyperkinesis are well known variables in psychological literature, and may play a part in the child's ability to become involved or to act effectively in different roles. Perhaps we can classify roles in terms of their suitability for impulsive behavior just as games have been similarly classified. Then again, there is the clinical dimension of "tolerance for unrealistic experiences." Does one role require more tolerance for fantasy than another?

We might ask whether completion or truncation have anything to do with "task persistence?" Is concentration of one sort akin to concentration of another? Consistency or inconsistency in the use of imaginary objects seems psychologically akin to nondistractibility, or at least to the ability to maintain the "as if" response set. Although a distinction is made between focus and consistency in the theatre, they may be part of a single related psychological variable.

Elaboration may be appropriate for use by psychologists as another means for measuring creativity. Does elaboration in drama relate to elaboration in words, art, or music? If it is true that non-kinetic persons make the most kinetic responses in projective fantasies, we might expect excellence in one area to be inversely related to excellence in another. Although unidimensional models are currently popular in psychology (cognitive, learning, etc.), personality theory suggests more multi-dimensional models. We might expect, therefore, to get an uneven profile with individuals excelling in one area but not in another.

In a preliminary view of some of our results, it looks as if involvement in drama and competence in drawing (tested and assessed independently) may be positively correlated among boys, but not among girls. Possibly females have a more differentiated expressive profile, whereas males exhibit a more global response to expressive possibilities. If so, that finding suggests the reverse of the situation typical in instrumental or achievement areas where males are usually more differentiated (have an uneven profile across competences), and females are less differentiated (Sutton-Smith & Roberts, 1964, pp. 13-37, & 105). In any case, the importance of a more general examination of differentiated competences across expressive areas (drama, music, art, dance, etc.) and instrumental areas (reading, arithmetic, science, etc.) and their relationships is certainly indicated.

Variables of space and bodily expression might have something to do with a person's sense of body barriers, or his body image, which are also well known psychological variables. At the very least, the deployment and expression of the body in dramatic space should permit studies of that sort (Editor's note: The dissertations of Bahs [1969] and Stockwell [1970] illustrate the fruitfulness of this line of inquiry).

Appropriateness of facial, verbal and gestural expression might also be related to other more comprehensive characteristics of expressive style. Perhaps social appropriateness is reflected in one's level of competence in social interaction, empathy, or social insight.

The most important point about these speculative associations, is that if it becomes possible to develop reliable techniques for assessing dramatic involvement using the measuring instrument described above, then it may provide us with a new means for exploring the function of many other psychological variables in human personality and development.

Some Dramatic Considerations

Regardless of the implications for psychology, the more important immediate yield must lie in what they can do for the understanding of child development through drama, and for the analysis of the theatrical experience, per se. Our immediate impression, for example, is that through the course of two months training, all of the children became more capable of focussing attention, of completion, of the consistent use of imaginary objects, of the relevant use of space, of elaboration, of projection and of appropriate facial, verbal, body use and social interaction. We must insist that this is a primary gain for drama. Because of their experience children will probably be more competent both in the performance and understanding of drama in subsequent years. More important, perhaps, is the notion that these children will also be more competent in the macrocosm of human socialization.

It is our expectation, furthermore, that this research will lead to subsequent studies in which the transfer of drama training effects into other areas of school life may be explored.

In our work with children in improvisational drama, an unmistakable

parallel seems to emerge between their developmental behavior and that of the adult professional actor in the process of creating a performance. Further, those uniquely discernable modes of behavior exhibited by certain children in informal dramatics seem characteristic of various styles and types of rehearsal patterns of professional actors. Other more sophisticated similarities have occurred all of which suggest a fruitful avenue of inquiry which tests the hypothesis that adult professional actors and children in creative dramatics go through the same or similar general processes. Perhaps one day it may be possible to predict accurately and easily dramatic aptitude or potential at various age levels.

REFERENCES

Dr. Gil Lazier is Associate Professor of Theatre in Research and Dramatic Theory at Florida State University. Dr. Brian Sutton-Smith is Professor of Psychology and Principal Advisor in Developmental Psychology at Teachers College, Columbia University. The authors gratefully acknowledge the support of the U. S. Department of Health, Education, and Welfare, TITLE IV (Project No. 90032), for research subsidy leading to concepts reported herein. Additional support for some of the research was received from the Public Health Service Grant # 5 ROI-MHI-7586-02. The students who took part in recording the behaviors of the drama teachers were: James Lee Austin, Rebecca Kaiser, and Neil Martin. The students who participated in developing the involvement scale were: Rosalind Barnett, Bob David, Steve Goldman, Birgitte Mednick, Amy Miller, and Mariam Viselman.

Bahs, C. H. The Effect of the Nature and Degree of Body-Cathexis on Pantomimic Movement. Unpublished Ph. D. dissertation, Bowling Green State University, 1969.

Berger, I. B. Creative Play Activities. New York: The Ronald Press, 1966.

Gottman, E. Behavior in Public Places. New York: Free Press, 1963.

Lazier, G. Dramatic Improvisation as English-Teaching Methodology, The English Record, 1969.

Lazier, G. N., Sutton-Smith, B. Assessment of Role Induction and Role Involvement in Creative Drama. First Progress Report, OE Bureau of Research No. 90032, U. S. Office of Education Bureau of Research, 1969.

Sarbin, T. Role Theory, Handbook of Social Psychology. ed. Lindsey G. Cambridge: Addison-Wesley, 1954.

Siks, G. B. Creative Dramatics: An Art for Children. New York: Harper and Row, 1958.

Spolin, V. Improvisation for the Theatre. Evanston, Illinois: Northwestern University Press, 1963.

Stanislavski, K. My Life in Art. New York: Theatre Arts, 1924.

Stanislavski, K. An Actor Prepares. New York: Theatre Arts, 1936.

Stanislavski, K. Building a Character. New York: Theatre Arts, 1949.

Stockwell, J. C. The Effects of the Body Image Dimensions of Satisfaction and Boundary on Pantomimic Movement. Unpublished Ph. D. dissertation, Bowling Green State University, 1970.

Sutton-Smith, B. Roberts, J. M. Rubrics of Competitive Behavior. Journal of Genetic Psychology. 1964, 105, 13-37.

Thurman, A., Bowers, N. D. The Development of Instruments Designed to Access Pupil Ability in Creative Dramatics. Paper presented at Childrens Theatre Conference, Los Angeles, California, 1968.

Ward, W. Playmaking with Children. New York: Appleton-Century-Crofts, 1957.

THE EFFECTIVENESS OF READER'S THEATRE AND SILENT READING IN COMMUNICATING THE MEANING OF UNFAMILIAR DRAMATIC LITERATURE

Howard D. Doll

Although the possibilities and needs for experimental research in oral interpretation have been recognized for some time, objective research has been sporadic and has not kept pace with research in other areas of speech (Clark, 1949; Green, 1949; Klyn, 1965; Brooks, 1967). The theses and dissertations devoted to experimental research in this area since 1949 number less than 15 (Brooks, 1967, p. 426). Certainly the area is ripe for further investigation, especially in view of the present renewed interest in Reader's Theatre.

In the experiments completed through 1969, the two major concerns seemed to be the nature of audience response to different methods of presentation and the relationship of oral to silent reading. The present study was designed to investigate the relationship between oral and silent reading in communicating the meaning of unfamiliar dramatic material. Specifically the purpose of this study was to investigate the effectiveness of Reader's Theatre and silent reading in communicating the meaning of The Bald Soprano to college students who had never read or seen the play before, and to determine what differences would pertain under various methods of presentation.

PROCEDURE

Subjects

Two-hundred and fifty-eight subjects enrolled in the 1969 fall quarter Fundamentals of Speech and Oral Interpretation classes at Ohio University were selected for participation in the experiment and were divided into two major populations: Population 1 was composed of 225 freshmen Fundamentals of Speech students (103 males and 122 females) randomly distributed into five separate groups for participation in preselected experimental conditions. Population 2 was composed of 33 sophomores, juniors, and seniors enrolled in Oral Interpretation. Owing to the large proportion of females (24), and the small N, subjects in Population 2 were subdivided into only two experimental groups.

Experimental Conditions

Subjects from Population 1 were assigned to five posttest experimental conditions: Group A silently read a mimeographed copy of The Bald Soprano after which they marked a set of semantic differentials. Group B observed a Reader's Theatre presentation of the same play following which they marked another identical set of semantic differentials. Immediately after reading the play, Group C saw a Reader's Theatre presentation after which they too marked the differentials. Group D saw a presentation of the play, read it silently, and then marked the differentials. Group E responded to the differentials twice. The first time they marked the differentials after having read the play silently, then, two weeks later they responded again after seeing a Reader's Theatre presentation of the play. The following chart simplifies the presentation of the experimental stimuli and the responses to them:

	Population 1 (Speech Fundamentals)	Population 2 (Oral Interpretation)
Group A	Silent reading followed by semantic differential	Silent reading followed by semantic differential
Group B	Reader's Theatre followed by semantic differential	Reader's Theatre followed by semantic differential
Group C	Silent reading, then Reader's Theatre followed by semantic differential	
Group D	Reader's Theatre, then silent reading followed by semantic differential	
Group E	Phase I: Silent reading followed by semantic differential Phase II: 2 weeks later. Reader's Theatre followed by semantic differential.	

The Stimulus

The selection of Eugene Ionesco's The Bald Soprano as a stimulus for the experiment was based on: (1) Ionesco's established reputation as a playwright of international fame, (2) the play's ability to be read silently or performed orally in one hour or less, and (3) the likelihood of the subject's unfamiliarity with the script.

A critical analysis of the script by the experimenter revealed eight themes or concepts each of which provided focus for the dependent measure. The concepts were: (1) the value of words, (2) the possibility of change, (3) love in marriage, (4) a typical life, (5) talking and listening, (6) the problems of others, (7) the sense of time, and (8) experience as a teacher.

The Dependent Measures

Eight identical copies of Osgood's semantic differential (Osgood, 1957, pp. 62-63) were used to measure the eight dramatic themes enumerated above. Each differential was composed of 18 scales each of which measured one of three factors: (1) An evaluative factor was measured by the scales: good-bad, optimistic-pessimistic, complete-incomplete, timely-untimely, positive-negative, and beautiful-ugly. (2) A potency factor was measured by: hard-soft, strong-weak, heavy-light, serious-humorous, and opaque-transparent. (3) An activity factor was measured by the scales: active-passive, complex-simple, hot-cold, excitable-calm, meaningful-meaningless, interesting-boring, and important-unimportant.

Administration of The Experiment

A pretest to measure the initial connotative meaning of the eight themes was administered by the experimenter two to four weeks prior to posttesting. The pretest was administered in a classroom setting three times a night for three nights with approximately 50 subjects assigned to each pretest section. To obscure the intent of the experiment and to discourage patterned responses, the polarity of the adjective pairs was reversed randomly, and the order of the scales was determined by random assignment.

The instrument used to obtain posttest measurements was identical to that used to obtain pretest scores; only the format (the order and polarity of scales) was changed.

The posttest was administered in two phases to Group E: (Phase I) two weeks before all of the other groups were posttested, Group E read The Bald Soprano silently. After all of the subjects had read the play they were asked to "mark the semantic differentials now that you have read the play." Observation of the data following the posttest indicated that no appreciable change had occured--later statistical tests (see Table 4) supported the validity of the observation. Hence in subsequent testing, subjects were given more specific instructions; they were told to "mark the semantic differentials in relation to your understanding of the play." (Phase II) two weeks following the initial posttest, Group E joined the other groups to receive their second posttest. The following chart provides a description of the physical conditions under which the stimuli and posttests were administered to the subjects:

Group	Location	Time	Activity
Group A	Copeland 111	7:00 PM 8:00 PM	Silent Reading Posttest
Group C	Bentley 240	7:00 PM	Silent Reading
Group B C D E	Bentley 240	8:00 PM	Reader's Theatre
Group B C E	Bentley 240	9:00 PM	Posttest
Group D	Copeland 111	9:00 PM 10:00 PM	Silent Reading Posttest

All of the activities described above were conducted in ordinary classrooms. Those subjects who were exposed to the Reader's Theatre presentation saw the same production. Those who read the play silently did so in a different classroom. After being exposed to the appropriate stimuli, the subjects were asked to respond to the posttest "in relation to their understanding of the play."

The experimenter was assisted by a graduate student in the conduct of the final part of the experiment.

Elimination of Subjects

Of the original 375 subjects who took part in the pretest, 117 were eliminated from final calculations. Ninty-seven were eliminated for failing to participate in the posttest, 11 because they had read the play previously, and 9 for failing to put their names on the answer sheets or marking the sheets improperly. Of the remaining 258 subjects, 82 were assigned to Group A, 54 to Group B, 42 to Group C, 45 to Group D, and 35 to Group E.

Analysis of Data

The data were analyzed by computer in four stages: (1) analysis of scale scores, (2) concept scores, (3) factor scores, and (4) group scores.

After mean scale scores for each subject had been computed, mean scores were computed for each factor on each concept, for all factors on each concept, and for each group and sex separately. Because of the polarity of the adjective pairs high mean scores should be interpreted as being indicative of favorable attitudes toward the concepts, and low mean scores as indicative of unfavorable attitudes.

RESULTS

Concept Scores

A comparison of the differences between the mean scale scores for all subjects for each concept on the pretest and the posttests revealed mean differences ranging from .84 to 1.40 on Posttest 1 and 1.02 to 1.69 on Posttest 2. Those mean scores and differences are displayed in Table 1. Although no statistical tests were applied to the means, the observed differences seemed of sufficient magnitude to warrant further analysis.

Factor Scores

The differences between the pretest scores and the Posttest 2 scores for the three factors (evaluation, potency, and activity) collapsed across all groups was tested statistically using the t statistic for non-homogenous groups (Lewis, 1960, p. 302). The results of those tests are displayed in Table 2 where significant differences were revealed for each of the three factors. It is apparent, then, that attitudes were affected by the conditions of the experiment. A more detailed exploration of the kinds and types of attitude change follows.

Group Scores

The mean scores by group, test, and sex, along with the differences between pretest and posttest scores by group and sex are displayed in Table 3. The Table indicates that, with respect to all of the groups, the difference between pretest and posttest means is greater for females than it is for males.

TABLE 1

DIFFERENCES BETWEEN MEAN SCORES ON PRETEST AND POSTTESTS FOR EACH CONCEPT

	Pretest	Post 1	Difference Pre-Post 1	Post 2	Difference Pre-Post 2
Value of Wds	4.79	3.83	.96	3.64	1.15
Poss of Chng	4.86	4.02	.84	3.84	1.02
Love In Mrrg	5.55	4.15	1.40	3.86	1.69
Typical Life	4.84	3.65	1.19	3.52	1.32
Talking & Lstng	5.00	3.86	1.14	3.71	1.29
Prblms of Othrs	4.72	3.85	.87	3.69	1.03
Sense of Time	4.63	3.49	1.14	3.31	1.32
Exp as Tchr	5.11	4.06	1.05	3.96	1.27

TABLE 2

THE MEANS, STANDARD DEVIATIONS, N'S, AND VALUES FOR t FOR THE THREE FACTORS ON THE PRETEST AND POSTTEST 2

	Pretest		Posttest 2			
Factor	Mean	σ	Mean	σ	t	N
Evaluation	4.97	.59	3.52	1.12	18.59*	258
Potency	4.48	.47	3.80	.82	11.53*	258
Activity	5.24	.59	3.75	1.24	16.39*	258

*Significant at .001 level.

When the differences were tested statistically (t-test), only Group C was associated with a significant difference (t = 2.70 at .01 level for d.f. = 40).

Values for t were then computed between the pretest and posttest for all of the groups. The results of those tests are displayed in Table 4. Statistically significant differences were revealed for Groups A (both populations), B (both populations), C, and E following the Reader's Theatre presentation. The differences for Group E following silent reading and for Group D were not found to be statistically significant at the .001 level of confidence.

The above tests indicate rather clearly that meaning is conveyed by exposure to either reading or observing a Reader's Theatre presentation of The Bald Soprano. Those tests do not, however, allow us any indication whether one type of exposure to literature is more effective in communicating meaning than another, and that, after all, is the major concern of the study.

In an attempt to determine which method is the most effective in its ability to communicate meaning, the t-statistic was used, once again, to test the differences between the posttest scores for each of the different groups.

TABLE 3

THE MEAN SCORES, N'S AND MEAN DIFFERENCES BETWEEN MEAN SCORES ON TESTS, USING DATA ON ALL THREE FACTORS

Group	N	Pretest	Posttest	Difference
GROUP A (ALL)	82	5.07	3.50	1.57
Male	30	4.96	3.66	1.30
Female	52	5.13	3.41	1.72
Population 1	66	5.04	3.53	1.52
Male	27	4.96	3.62	1.34
Female	39	5.09	3.45	1.64
Population 2	16	5.22	3.43	1.79
Male	3	4.97	3.88	1.09
Female	13	5.28	3.28	2.00
GROUP B (ALL)	54	5.01	3.47	1.64
Male	23	4.93	3.53	1.40
Female	31	5.06	3.43	1.63
Population 1	37	4.95	3.43	1.52
Male	17	4.85	3.46	1.39
Female	20	5.05	3.41	1.64
Population 2	17	5.12	3.56	1.56
Male	6	5.18	3.73	1.45
Female	11	5.08	3.46	1.62
GROUP C	42	4.84	3.36	1.48
Male	21	4.75	3.74	1.01
Female	21	4.93	2.97	1.96
GROUP D	45	4.86	4.85	.01
Male	22	4.71	4.85	.14 +
Female	23	4.99	4.84	.15
GROUP E (PHASE I)	35	4.74	4.65	.09
Male	16	4.83	4.79	.04
Female	19	4.67	4.53	.14
(PHASE II)	35	4.74	3.37	1.37
Male	16	4.83	3.45	1.38
Female	19	4.67	3.31	1.37

+Difference is in opposite direction from all other differences.

The results of those tests may be found in Table 5.

It is apparent from Table 5 that Groups D and E (Phase I) do not differ significantly from each other, but that, together, they differ from all of the other groups. Likewise, the other groups (A, B, C, and E [Phase II]), though differing significantly from Groups D and E (Phase I), do not differ significantly from each other. It is apparent from the group mean scores, Table 4, that, although the subjects in Groups A, B, C, and E (Phase II) experienced a significant shift in meaning following exposure to the experimental stimulus (the play), the subjects in Groups D and E (Phase I) did not experience a similar shift. For

TABLE 4

VALUES FOR t FOR DIFFERENCES IN MEAN SCORES BETWEEN THE PRETEST AND POSTTEST FOR EACH GROUP FOR ALL FACTORS

Group	d.f.	t	Group	d.f.	t
GROUP A	80	13.77*	GROUP C	40	9.13*
Population 1	66	11.92*			
Population 2	16	6.88*	GROUP D	43	.08
GROUP B	52	12.01*	GROUP E		
Population 1	35	10.33*	Phase I	33	.61
Population 2	15	6.19*	Phase II	33	8.30*

*Significant at the .001 level.

TABLE 5

VALUES FOR t FOR DIFFERENCES BETWEEN GROUP MEAN POSTTEST SCORES

	GROUP B	GROUP C	GROUP D	GROUP E (I)	GROUP E (II)
GROUP A	.18	.80	8.55*	6.40*	.71
GROUP B		.64	9.09*	6.80*	.57
GROUP C			8.63*	6.54*	.60
GROUP D				1.26	8.79*
GROUP E (PHASE I)					6.63*

*Significant at the .001 level.

Group E (Phase I), at least, the failure to detect such a shift may be easily explained as resulting from the lack of specificity in the instructions given during the administration of the posttest (see p.49). The idiosyncratic results obtained from the subjects in Group D are not so easily explained, however, and suggest the need for further research into the effects of the relationships of the two different methods of perceiving the play.

As a conclusion to the experiment, the differences between the two major populations were tested using the t-test once again. The results are displayed in Table 6 and reveal no significant differences between the populations (Population 1, and 2).

CONCLUSIONS

1. Unless subjects were told specifically to look for meaning following the reading or performance of the play, there was no shift in meaning from pretest to posttest.

TABLE 6

VALUES FOR t FOR DIFFERENCES BETWEEN POPULATION MEAN POSTTEST SCORES

	GROUP A Population 2	GROUP B Population 1	GROUP B Population 2
GROUP A			
Population 1	.34	.47	.16
Population 2		.02	.39
GROUP B			
Population 1			.52

*Significant at the .001 level.

2. When subjects were told to look for meaning, both methods of presentation--silent reading and Reader's Theatre--were found to communicate the meaning of unfamiliar dramatic literature to college students effectively.

3. The commonly held notion that Reader's Theatre is a more effective device for conveying the meaning of unfamiliar material was not supported by the findings of this study; both methods were found to be equally effective.

4. The study also failed to support the superiority of combined methods of presentation. As a matter of fact, when the Reader's Theatre presentation was followed by silent reading, meaning did not appear to be conveyed to the subjects effectively.

5. Although females generally and consistently held more positive initial attitudes than males, and though the shift in meaning was greater for the females, the difference was statistically significant for only one group.

6. Silent reading alone and Reader's Theatre alone communicated the meaning of unfamiliar literature equally and as well as any combination of the two methods.

REFERENCES

Dr. Howard Doll is presently an Assistant Professor of Speech at the University of North Carolina, Chapel Hill. This article is based on a doctoral dissertation completed at Ohio University under the direction of Dr. Carl H. Weaver, Professor of Speech and Co-ordinator of Graduate Studies in the School of Interpersonal Communications, Ohio University.

Brooks, K., Bahn, E., and Okey, L. L. The Communicative Act of Oral Interpretation. Boston: Allyn and Bacon, Inc., 1967, pp. 425-27.

Clark, R. D. Research Possibilities in Interpretative Speech. Western Speech Journal, 1949, 12, 16-19.

Green, C. P. Graduate Study in Oral Interpretation. Southern Speech Journal, 1949, 15, 128-37.

Klyn, M. S. Potentials for Research in Oral Interpretation. Western Speech Journal, 1965, 29, 108-13.

Lewis, D. Quantitative Methods in Psychology. New York: McGraw-Hill Book Co., 1960, p. 302.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. The Measurement of Meaning. Urbana: The University of Illinois Press, 1957, pp. 62-3.

PERSONALITY AND CONTROVERSIAL MATERIAL ON THE STAGE

Christopher M. Reynolds

Theatre producers and directors are required sometimes to make decisions about what might or might not offend their audiences. Unfortunately such decisions are supported by little or no empirical evidence regarding probable effects. In an effort to expand our understanding of the effects of different kinds of controversial stage material on various members of an audience, this study was conducted.

Specifically the study investigated the relationship of selected personality variables within members of an audience to the degree to which those members might be expected to either approve or disapprove of theatrical scenes involving such things as sex, nudity, violence, or other controversial material.

To implement the study two different sets of data were collected from a single group of 100 selected college students. Correlation coefficients were then obtained from the two sets of data in an attempt to reveal the degree of the relationship between the different personality traits and the degree of approval or disapproval.

To measure the latter, the Reynolds Theatre Approval Scale was used (although the instrument is protected by copywright, copies may be obtained without charge by request from the author; Christopher M. Reynolds, Department of English, Winthrop College, Rock Hill, South Carolina, 29730). The instrument was developed, tested, and validated as a part of the author's doctoral dissertation (Reynolds, 1969) and includes responses to eight different dimensions of controversial stage material: (1) the use of drugs or ahcohol, (2) the use of questionable language, (3) the use of revealing costumes or nudity, (4) questionable presentation of characters associated with some phase of religion, (5) the display of sexual activity, (6) the display of acts of violence, (7) whether plays could function as harmful influences, (8) whether the testee would wish to take action against plays he found objectionable.

In order to obtain the personality data the following standard psychological tests were administered: (1) the California F Scale, (2) the California Psychological Inventory, (3) McClosky's Conservatism Scale, (4) Rokeach's D Scale, and (5) the Runner Studies of Attitude Patterns. From these instruments 26 personality traits were extracted in an attempt to reveal the following five basic personality patterns: (1) authoritarianism, (2) conservatism, (3) hostility, (4) practicality and conventionality, and (5) creative orientation.

The coefficients describing the relationships of the two sets of data are displayed in the Table. Those findings suggest that the disapproving (or repressive) responder to controversial theatre material tended to be more authoritarian, hostile, conservative, practical, and conventional, whereas the more approving responder tended to be more non-authoritarian, generally more liberal, lacking in planful practicality, and more unconventional.

TABLE

COEFFICIENTS OF CORRELATION BETWEEN THE REYNOLDS TEST AND PERSONALITY TESTS

Reynolds Test +			
Positive Correlations		Negative Correlations	
Flexibility	.59**	Rules Oriented	-.57**
Freedom Oriented	.27**	F Scale	-.57**
Intellectual Efficiency	.25*	Classical Conservatism	-.54**
Intuitive Orientation	.22*	Control Oriented	-.50**
Experimental Orientation	.20*	D Scale	-.45**
Resistance to Social Pressure	.18	Planful Practicality	-.44**
Dominance	.09	Hostility	-.39**
Extraversiveness	.08	Performance Anxiety	-.28**
Power Drive	.03	Communality	-.27**
		Anxiety Oriented	-.26**
		Femininity	-.23*
		Anomie	-.23*
		Social Anxiety	-.18
		Passive Compliance	-.14

+ The Reynolds Test was scored so that a high score was indicative of approval. Hence negative correlations suggest that the personality trait is negatively correlated with approval and positive correlations indicate positive relationships.

* Significant at .05 level of confidence.

**Significant at .01 level of confidence.

While one should exercise caution when generalizing to a general population (because of the size and breadth of the sample population), the study suggests that resistance toward controversial subject material on stage is but another facet of an individual's central pattern of rigidity, hostility, and rejection of change.

Reference

Dr. Reynolds is Associate Professor of Drama at Winthrop College, Rock Hill, South Carolina. This paper is based on a Ph. D. Dissertation completed at the University of Michigan in 1969, under the direction of Kenneth E. Andersen. Special assistance in psychological testing was given by William G. Murdy, Department of Psychology, Winthrop College.

Reynolds, C. M. The Reynolds Test: Theatre Approval Scale. Unpublished Ph. D. Dissertation. Ann Arbor: University of Michigan, 1969.

STYLE AND TEMPERAMENT

Warren Pickett

That style is a function of the temperament of the artist and as such conveys an impression of the artist's temperament to his audience is a commonplace assumption among theatre literati. With almost equal frequency it is assumed as a corollary that the appeal of a particular style is singular among individual audience members. Simply, one style may appeal more to one audience member than to another. Despite the regularity with which one encounters both the parent theory and its corollary hypothesis, neither has been subjected to the rigor of empirical testing. This study makes an initial attempt to satisfy that need by testing the hypothesis that there is more than a chance relationship between a particular style and the temperament of the audience member to whom it appeals. To be more specific, the study is an empirical investigation of the relationship between three set design styles--Romantic, Formalistic, and Baroque--to two temperament types: introversion and extroversion.

The subject population was composed of 300 entering freshmen at Windham College, Vermont, who were dichotomized into two experimental groups--introverted and extroverted--on the basis of the way in which they scored on the "Introversion-Extroversion Scale" of the Bernreuter Personality Inventory. Using the 98th percentile on either end of the scale as cut-off points, two experimental groups, each containing 13 males and 7 females, were formed.

To provide assurance that the subjects in neither group had been drawn from different parent populations, a check was made of the subjects' records at the College Admissions Office, and a questionnaire was administered to all of the participants. Neither source of information provided grounds for suspecting that the subjects from either group differed substantially with regard to relevant variables other than those crucial to the experiment. The subjects ranged in age from 17 to 21, and they scored from 400 to 600 on their College Board Exams. Most had had very little formal instruction in the arts, and very little or no exposure to the ligitimate stage or to art museums. More than 95% of the subjects had come from families with an average annual income of approximately $30,000, and from families, 95% of which, were located within the greater metropolitan New York City area. Their answers suggested further that there were no professional artists in their families nor among their close friends, and that only a few among their families or friends had even an avocational interest in any of the arts.

Romantic, Formalistic, and Baroque styles were chosen as stimuli because the first two were considered by many authorities to be representative, respectively, of introverted and extroverted temperaments, while the third, the Baroque, is often cited as a style representative of a temperament falling somewhat between those two extremes. Although all three styles are generally associated with distinct historical periods (the Romantic with the first half of the 19th century, Formalism with the first half of the 20th century, and Baroque

with the latter half of the 17th and early part of the 18th centuries), all are sufficiently generic to find representative examples throughout the entire history of art.

Ten examples of each of the three styles were selected by the investigator from among the works of reputable historians. Each example was a black and white drawing by the designer of a set from a period which had been designated, by the historian in whose work the drawing was found, as representative of one of the three styles. Since Rohrshach protocols had indicated that extroverts have a very stong praference for color, colored drawings were not included in the experimental samples. To further avoid the risk of contamination by extraneous variables, drawings which included human or animal figures were excluded also. To simplify handling, and to allow for uniformity among the size of the samples, all of the drawings were photographically reduced to standard postcard size.

As a further validity check on the degree to which the examples were truly typical of the style they were intended to represent, all of them were judged by five members of the faculty of the University of Michigan. Four of the five were members of the theatre faculty, while the fifth, an experimental psychologist, had been employed professionally in the legitimate theatre. After the order of the drawings had been randomized for style by the investigator, they were presented to the faculty judges who were instructed: (1) to sort the drawings into three piles, each representative of one of the three experimental styles, and then (2) to rank each of the sorted drawings according to the degree to which that drawing was typical of the style to which it had been assigned. On the basis of those rankings, four drawings from each of the three styles (12 in all) were ultimately selected for experimental use.

The 12 drawings were then presented to each of the 40 subjects individually. The subjects were asked to rate each of the drawings according to preference on a seven interval continuum, with each interval further subdivided into quarters allowing a total of 25 rating options. The subjects were told that a score of "7" should be reserved for their highest preference and a score of "1" for their lowest.

The following is a summary of the preferences:

	Introverts	Extroverts
Highest Preference	Baroque	Romantic
	Formal	Baroque
Lowest Preference	Romantic	Formal

Statistical tests applied to the data suggested that the differences in preference were statistically significant beyond the .05 level. Style preferences were also correlated with some of the background information obtained from the subjects such as age, sex, and intelligence, but none of the coefficients of correlation was found to be statistically significant.

Although the results of the study confirm the major hypothesis (that

introverts will have different style preferences than extroverts), the preferences were not in the predicted directions. The introverts did not prefer the Romantic, nor did the extroverts prefer the Formal, nor was the Baroque the intermediate choice of both temperaments.

In many respects the study raises more questions than it answers. According to most of the authorities, Romanticism should have been the clear choice for the introverts since it tends to emphasize the private emotional life of the individual along with a withdrawal from society with all of its demands, restrictions, and mechanical trappings. We think of all of those things as alien to the introvert. How then does one account for the introvert's seeming preference for the comparatively rigid, restricted, demanding, and aggressive style of the Baroque? Does the study cast doubt on the traditional descriptions of the style or of the temperament, or is it possible that the introvert, aware of an imbalance in his temperament, seeks compensation in a style which runs against the grain of his natural temperament?

Whatever the case the study opens the door to some very interesting research possibilities. Certainly the study has not exhausted the many potential style temperament relationships nor has it exposed the causative links between style and temperament. What it has done is to provide hard evidence that such links do exist and that style preference is not monolithic but varies according to the temperament of the viewer. It further points to the possibility that further research to expose these links may not only provide the director with a more effective technique but may also lead to an understanding of the nature and function of style; an understanding based upon evidence rather than intuition.

Reference

Dr. Warren Pickett is currently teaching at the Colegio Internacional, Baleares, Spain. This article is abstracted from his dissertation, cited below.

Pickett, Warren An Experiment in the Preference of Different Temperament Types for Different Historical Styles of Set Design. Unpublished Ph. D. Dissertation. Ann Arbor: University of Michigan, 1969.